BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

As rescanning documents *will not* correct images, please do not report the images to the Image Problem Mailbox.

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
19 September 2002 (19.09.2002)

PCT

(10) International Publication Number
WO 02/072794 A2

(51) International Patent Classification[7]: C12N

(21) International Application Number: PCT/US02/09052

(22) International Filing Date: 12 March 2002 (12.03.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

60/275,249	12 March 2001 (12.03.2001)	US
60/316,810	31 August 2001 (31.08.2001)	US
60/323,977	21 September 2001 (21.09.2001)	US
60/348,447	26 October 2001 (26.10.2001)	US
60/343,880	2 November 2001 (02.11.2001)	US

(71) Applicant *(for all designated States except US)*: INCYTE GENOMICS, INC. [US/US]; 3160 Porter Drive, Palo Alto, CA 94304 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: YUE, Henry [US/US]; 826 Lois Avenue, Sunnyvale, CA 94087 (US). XU, Yuming [US/US]; 1739 Walnut Drive, Mountain View, CA 94040 (US). THANGAVELU, Kavitha [IN/US]; 1950 Montecito Avenue #23, Mountain View, CA 94043 (US). WARREN, Bridget, A. [US/US]; 10130 Parkwood Drive # 2, Cupertino, CA 95014 (US). TANG, Y., Tom [US/US]; 4230 Ranwick Court, San Jose, CA 95118 (US). DUGGAN, Brendan, M. [AU/US]; 243 Buena Vista Avenue # 306, Sunnyvale, CA 94086 (US). TRAN, Uyen, K. [US/US]; 2638 Mabury Square, San Jose, CA 95133 (US). BAUGHN, Mariah, R. [US/US]; 14244 Santiago Road, San Leandro, CA 94577 (US). HONCHELL, Cynthia, D. [US/US]; 400 Laurel Street #203, San Carlos, CA 94070 (US). BURFORD, Neil [GB/US]; 105 Wildwood Circle, Durham, CT 06422 (US). FORSYTHE, Ian, J. [US/US]; 308 Roble Avenue, Redwood City, CA 94061 (US). YANG, Junming [CN/US]; 7125 Bark Lane, San Jose, CA 95129 (US). MASON, Patricia, M. [US/US]; 360 Clarke Lane, Morgan Hill, CA 95037 (US).

(74) Agents: HAMLET-COX, Diana et al.; Incyte Genomics, Inc., 3160 Porter Drive, Palo Alto, CA 94304 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZM, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 02/072794 A2

(54) Title: IMMUNOGLOBULIN SUPERFAMILY PROTEINS

(57) Abstract: The invention provides human immunoglobulin superfamily proteins (IGSFP) and polynucleotides which identify and encode IGSFP. The invention also provides expression vectors, host cells, antibodies, agonists, and antagonists. The invention also provides methods for diagnosing, treating, or preventing disorders associated with aberrant expression of IGSFP.

IMMUNOGLOBULIN SUPERFAMILY PROTEINS

TECHNICAL FIELD

This invention relates to nucleic acid and amino acid sequences of immunoglobulin superfamily proteins and to the use of these sequences in the diagnosis, treatment, and prevention of immune system, neurological, developmental, muscle, and cell proliferative disorders, and in the assessment of the effects of exogenous compounds on the expression of nucleic acid and amino acid sequences of immunoglobulin superfamily proteins.

BACKGROUND OF THE INVENTION

Most cell surface and soluble molecules that mediate functions such as recognition, adhesion or binding have evolved from a common evolutionary precursor (i.e., these proteins have structural homology). A number of molecules outside the immune system that have similar functions are also derived from this same evolutionary precursor. These molecules are classified as members of the immunoglobulin (Ig) superfamily. The criteria for a protein to be a member of the Ig superfamily is to have one or more Ig domains, which are regions of 70-110 amino acid residues in length homologous to either Ig variable-like (V) or Ig constant-like (C) domains. Members of the Ig superfamily include antibodies (Ab), T cell receptors (TCRs), class I and II major histocompatibility (MHC) proteins, CD2, CD3, CD4, CD8, poly-Ig receptors, Fc receptors, neural cell-adhesion molecule (NCAM) and platelet-derived growth factor receptor (PDGFR).

Ig domains (V and C) are regions of conserved amino acid residues that give a polypeptide a globular tertiary structure called an immunoglobulin (or antibody) fold, which consists of two approximately parallel layers of β-sheets. Conserved cysteine residues form an intrachain disulfide-bonded loop, 55-75 amino acid residues in length, which connects the two layers of the β-sheets. Each β-sheet has three or four anti-parallel β-strands of 5-10 amino acid residues. Hydrophobic and hydrophilic interactions of amino acid residues within the β-strands stabilize the Ig fold (hydrophobic on inward facing amino acid residues and hydrophilic on the amino acid residues in the outward facing portion of the strands). A V domain consists of a longer polypeptide than a C domain, with an additional pair of β-strands in the Ig fold.

A consistent feature of Ig superfamily genes is that each sequence of an Ig domain is encoded by a single exon. It is possible that the superfamily evolved from a gene coding for a single Ig domain involved in mediating cell-cell interactions. New members of the superfamily then arose by exon and gene duplications. Modern Ig superfamily proteins contain different numbers of V and/or C

domains. Another evolutionary feature of this superfamily is the ability to undergo DNA rearrangements, a unique feature retained by the antigen receptor members of the family.

Many members of the Ig superfamily are integral plasma membrane proteins with extracellular Ig domains. The hydrophobic amino acid residues of their transmembrane domains and their cytoplasmic tails are very diverse, with little or no homology among Ig family members or to known signal-transducing structures. There are exceptions to this general superfamily description. For example, the cytoplasmic tail of PDGFR has tyrosine kinase activity. In addition Thy-1 is a glycoprotein found on thymocytes and T cells. This protein has no cytoplasmic tail, but is instead attached to the plasma membrane by a covalent glycophosphatidylinositol linkage.

Another common feature of many Ig superfamily proteins is the interactions between Ig domains which are essential for the function of these molecules. Interactions between Ig domains of a multimeric protein can be either homophilic or heterophilic (i.e., between the same or different Ig domains). Antibodies are multimeric proteins which have both homophilic and heterophilic interactions between Ig domains. Pairing of constant regions of heavy chains forms the Fc region of an antibody and pairing of variable regions of light and heavy chains form the antigen binding site of an antibody. Heterophilic interactions also occur between Ig domains of different molecules. These interactions provide adhesion between cells for significant cell-cell interactions in the immune system and in the developing and mature nervous system. (Reviewed in Abbas, A.K. et al. (1991) Cellular and Molecular Immunology, W.B. Saunders Company, Philadelphia, PA, pp.142-145.)

Antibodies

Antibodies are multimeric members of the Ig superfamily which are either expressed on the surface of B-cells or secreted by B-cells into the circulation. Antibodies bind and neutralize foreign antigens in the blood and other extracellular fluids. The prototypical antibody is a tetramer consisting of two identical heavy polypeptide chains (H-chains) and two identical light polypeptide chains (L-chains) interlinked by disulfide bonds. This arrangement confers the characteristic Y-shape to antibody molecules. Antibodies are classified based on their H-chain composition. The five antibody classes, IgA, IgD, IgE, IgG and IgM, are defined by the α, δ, ε, γ, and μ H-chain types. There are two types of L-chains, κ and λ, either of which may associate as a pair with any H-chain pair. IgG, the most common class of antibody found in the circulation, is tetrameric, while the other classes of antibodies are generally variants or multimers of this basic structure.

H-chains and L-chains each contain an N-terminal variable region and a C-terminal constant region. The constant region consists of about 110 amino acids in L-chains and about 330 or 440 amino acids in H-chains. The amino acid sequence of the constant region is nearly identical among H- or L-

chains of a particular class. The variable region consists of about 110 amino acids in both H- and L-chains. However, the amino acid sequence of the variable region differs among H- or L-chains of a particular class. Within each H- or L-chain variable region are three hypervariable regions of extensive sequence diversity, each consisting of about 5 to 10 amino acids. In the antibody molecule, the H- and L-chain hypervariable regions come together to form the antigen recognition site. (Reviewed in Alberts, B. et al. (1994) Molecular Biology of the Cell, Garland Publishing, New York, NY, pp. 1206-1213 and 1216-1217.)

Both H-chains and L-chains contain the repeated Ig domains of members of the Ig superfamily. For example, a typical H-chain contains four Ig domains, three of which occur within the constant region and one of which occurs within the variable region and contributes to the formation of the antigen recognition site. Likewise, a typical L-chain contains two Ig domains, one of which occurs within the constant region and one of which occurs within the variable region.

The immune system is capable of recognizing and responding to any foreign molecule that enters the body. Therefore, the immune system must be armed with a full repertoire of antibodies against all potential antigens. Such antibody diversity is generated by somatic rearrangement of gene segments encoding variable and constant regions. These gene segments are joined together by site-specific recombination which occurs between highly conserved DNA sequences that flank each gene segment. Because there are hundreds of different gene segments, millions of unique genes can be generated combinatorially. In addition, imprecise joining of these segments and an unusually high rate of somatic mutation within these segments further contribute to the generation of a diverse antibody population.

Neural Cell Adhesion Proteins

Neural cell adhesion proteins (NCAPs) play roles in the establishment of neural networks during development and regeneration of the nervous system (Uyemura et al. (1996) Essays Biochem. 31:37-48; Brummendorf and Rathjen (1996) Curr. Opin. Neurobiol. 6:584-593). NCAP participates in neuronal cell migration, cell adhesion, neurite outgrowth, axonal fasciculation, pathfinding, synaptic target-recognition, synaptic formation, myelination and regeneration. NCAPs are expressed on the surfaces of neurons associated with learning and memory. Mutations in genes encoding NCAPS are linked with neurological diseases, including Charcot-Marie-Tooth disease (a hereditary neuropathy), Dejerine-Sottas disease, X-linked hydrocephalus, MASA syndrome (mental retardation, aphasia, shuffling gait and adducted thumbs), and spastic paraplegia type I. In some cases, expression of NCAP is not restricted to the nervous system. L1, for example, is expressed in melanoma cells and hematopoietic tumor cells where it is implicated in cell spreading and migration, and may play a role in

tumor progression (Montgomery et al. (1996) J. Cell Biol. 132:475-485).

NCAPs have at least one immunoglobulin constant or variable domain (Uyemura et al., supra). They are generally linked to the plasma membrane through a transmembrane domain and/or a glycosyl-phosphatidylinositol (GPI) anchor. The GPI linkage can be cleaved by GPI phospholipase C. Most NCAPs consist of an extracellular region made up of one or more immunoglobulin domains, a membrane spanning domain, and an intracellular region. Many NCAPs contain post-translational modifications including covalently attached oligosaccharide, glucuronic acid, and sulfate. NCAPs fall into three subgroups: simple-type, complex-type, and mixed-type. Simple-type NCAPs contain one or more variable or constant immunoglobulin domains, but lack other types of domains. Members of the simple-type subgroup include Schwann cell myelin protein (SMP), limbic system-associated membrane protein (LAMP) and opiate-binding cell-adhesion molecule (OBCAM). The complex-type NCAPs contain fibronectin type III domains in addition to the immunoglobulin domains. The complex-type subgroup includes neural cell-adhesion molecule (NCAM), axonin-1, F11, Bravo, and L1. Mixed-type NCAPs contain a combination of immunoglobulin domains and other motifs such as tyrosine kinase, epidermal growth factor-like, sema, and PSI (plexins, semaphorins, and integrins) domains. This subgroup includes Trk receptors of nerve growth factors such as nerve growth factor (NGF) and neurotropin 4 (NT4), Neu differentiation factors such as glial growth factor II (GGFII) and acetylcholine receptor-inducing factor (ARIA), the semaphorin/collapsin family such as semaphorin B and collapsin, and receptors for members of the semaphorin/collapsin family such as plexin (for plexin, see below).

An NCAP subfamily, the NCAP-LON subgroup, includes cell adhesion proteins expressed on distinct subpopulations of brain neurons. Members of the NCAP-LON subgroup possess three immunoglobulin domains and bind to cell membranes through GPI anchors. Kilon (a kindred of NCAP-LON), for example, is expressed in the brain cerebral cortex and hippocampus (Funatsu et al. (1999) J. Biol. Chem. 274:8224-8230). Immunostaining localizes Kilon to the dendrites and soma of pyramidal neurons. Kilon has three C2 type immunoglobulin-like domains, six predicted glycosylation sites, and a GPI anchor. Expression of Kilon is developmentally regulated. It is expressed at higher levels in adult brain in comparison to embryonic and early postnatal brains. Confocal microscopy shows the presence of Kilon in dendrites of hypothalamic magnocellular neurons secreting neuropeptides, oxytocin, or arginine vasopressin (Miyata et al. (2000) J. Comp. Neurol. 424:74-85). Arginine vasopressin regulates body fluid homeostasis, extracellular osmolarity and intravascular volume. Oxytocin induces contractions of uterine smooth muscle during child birth and of myoepithelial cells in mammary glands during lactation. In magnocellular neurons, Kilon is proposed to

play roles in the reorganization of dendritic connections during neuropeptide secretion.

Sidekick (SDK) is a member of the NCAP family. The extracellular region of SDK contains six immunoglobulin domains and thirteen fibronectin type III domains. SDK is involved in cell-cell interaction during eye development in Drosophila (Nguyen, D. N. T. et al. (1997) Development 124: 3303).

Synaptic Membrane Glycoproteins

Specialized cell junctions can occur at points of cell-cell contact. Among these cell junctions are communicating junctions which mediate the passage of chemical and electrical signals between cells. In the central nervous system, communicating junctions between neurons are known as synaptic junctions. They are composed of the membranes and cytoskeletons of the pre- and post-synaptic neurons. Some glycoproteins, found in biochemically isolated synaptic subfractions such as the synaptic membrane (SM) and postsynaptic density (PSD) fractions, have been identified and their functions established. An example is the SM glycoprotein, gp50, identified as the β2 subunit of the Na^+/K^+-ATPase.

Two glycoproteins, gp65 and gp55, are major components of synaptic membranes prepared from rat forebrain. They are members of the Ig superfamily containing three and two Ig domains, respectively. As members of the Ig superfamily, it is proposed that a possible function of these proteins is to mediate adhesive interactions at the synaptic junction. (Langnaese, K. et al. (1997) J. Biol. Chem.272:821-827.)

Lectins

Lectins comprise a ubiquitous family of extracellular glycoproteins which bind cell surface carbohydrates specifically and reversibly, resulting in the agglutination of cells (reviewed in Drickamer, K. and Taylor, M. E. (1993) Annu. Rev. Cell Biol. 9:237-264). This function is particularly important for activation of the immune response. Lectins mediate the agglutination and mitogenic stimulation of lymphocytes at sites of inflammation (Lasky, L. A. (1991) J. Cell. Biochem. 45:139-146; Paietta, E. et al. (1989) J. Immunol. 143:2850-2857).

Sialic acid binding Ig-like lectins (SIGLECs) are members of the Ig superfamily that bind to sialic acids in glycoproteins and glycolipids. SIGLECs include sialoadhesin, CD22, CD33, myelin-associated glycoprotein (MAG), SIGLEC-5, SIGLEC-6, SIGLEC-7, and SIGLEC-8. The extracellular region of SIGLEC has a membrane distal V-set domain followed by varying numbers of C2-set domains. The sialic acid binding domain is mapped to the V-set domain. Except for MAG which is expressed exclusively in the nervous system, most SIGLECs are expressed on distinct subsets of hemopoietic cells. For example, SIGLEC-8 is expressed exclusively in eosinophils, one form of

polymorphonuclear leucocyte (granulocyte) (Floyd, H. et al. (2000) J. Biol. Chem. 275: 861-866).

Leucine-Rich Repeat Proteins

Leucine-rich repeat proteins (LRRPs) are involved in protein-protein interactions. LRRPs such as mammalian neuronal leucine-rich repeat proteins (NLLR-1 and NLLR-2), Drosophila connectin, slit, chaopin, and toll all play roles in neuronal development. The extracellular region of LRRPs contains varying numbers of leucine-rich repeats, immunoglobulin-like domains, and fibronectin type III domains (Taguchi, A. et al. (1996) Brain Res. Mol. Brain Res. 35:31-40).

In addition to the V and C2 sets of immunoglobulin-like domains, there is a D set immunoglobulin-like domain, named IPT/TIG (for immunoglobulin-like fold shared by plexins and transcription factors). IPT/TIG containing proteins include plexins, MET/ RON/ SEA (hepatocyte growth factor receptor family), and the transcription factor XCoe2, a transcription factor of the Col/Olf-1/EBF family involved in the specification of primary neurons in Xenopus (Bork, P. et al. (1999) Trends in Biochem. 24:261-263; Santoro, N. M. et al. (1996) Mol. Cell Biol. 16:7072-7083; Dubois L. et al. (1998) Curr. Biol.8:199-209). Plexins such as plexin A and VESPR have been shown to be neuronal semaphorin receptors that control axon guidance (Winberg M. L. et al. (1998) Cell 95:903-916).

Expression profiling

Array technology can provide a simple way to explore the expression of a single polymorphic gene or the expression profile of a large number of related or unrelated genes. When the expression of a single gene is examined, arrays are employed to detect the expression of a specific gene or its variants. When an expression profile is examined, arrays provide a platform for identifying genes that are tissue specific, are affected by a substance being tested in a toxicology assay, are part of a signaling cascade, carry out housekeeping functions, or are specifically related to a particular genetic predisposition, condition, disease, or disorder.

The discovery of new immunoglobulin superfamily proteins, and the polynucleotides encoding them, satisfies a need in the art by providing new compositions which are useful in the diagnosis, prevention, and treatment of immune system, neurological, developmental, muscle, and cell proliferative disorders, and in the assessment of the effects of exogenous compounds on the expression of nucleic acid and amino acid sequences of immunoglobulin superfamily proteins.

SUMMARY OF THE INVENTION

The invention features purified polypeptides, immunoglobulin superfamily proteins, referred to collectively as "IGSFP" and individually as "IGSFP-1," "IGSFP-2," "IGSFP-3," "IGSFP-4," "IGSFP-

5," "IGSFP-6," "IGSFP-7," "IGSFP-8," "IGSFP-9," "IGSFP-10," "IGSFP-11," and "IGSFP-12." In one aspect, the invention provides an isolated polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-12, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-12, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-12, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-12. In one alternative, the invention provides an isolated polypeptide comprising the amino acid sequence of SEQ ID NO:1-12.

The invention further provides an isolated polynucleotide encoding a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-12, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-12, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-12, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-12. In one alternative, the polynucleotide encodes a polypeptide selected from the group consisting of SEQ ID NO:1-12. In another alternative, the polynucleotide is selected from the group consisting of SEQ ID NO:13-24.

Additionally, the invention provides a recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide encoding a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-12, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-12, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-12, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-12. In one alternative, the invention provides a cell transformed with the recombinant polynucleotide. In another alternative, the invention provides a transgenic organism comprising the recombinant polynucleotide.

The invention also provides a method for producing a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-12, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-12, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group

consisting of SEQ ID NO:1-12, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-12. The method comprises a) culturing a cell under conditions suitable for expression of the polypeptide, wherein said cell is transformed with a recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide encoding the polypeptide, and b) recovering the polypeptide so expressed.

Additionally, the invention provides an isolated antibody which specifically binds to a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-12, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-12, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-12, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-12.

The invention further provides an isolated polynucleotide selected from the group consisting of a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:13-24, b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:13-24, c) a polynucleotide complementary to the polynucleotide of a), d) a polynucleotide complementary to the polynucleotide of b), and e) an RNA equivalent of a)-d). In one alternative, the polynucleotide comprises at least 60 contiguous nucleotides.

Additionally, the invention provides a method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide selected from the group consisting of a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:13-24, b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:13-24, c) a polynucleotide complementary to the polynucleotide of a), d) a polynucleotide complementary to the polynucleotide of b), and e) an RNA equivalent of a)-d). The method comprises a) hybridizing the sample with a probe comprising at least 20 contiguous nucleotides comprising a sequence complementary to said target polynucleotide in the sample, and which probe specifically hybridizes to said target polynucleotide, under conditions whereby a hybridization complex is formed between said probe and said target polynucleotide or fragments thereof, and b) detecting the presence or absence of said hybridization complex, and optionally, if present, the amount thereof. In one alternative, the probe comprises at least 60 contiguous nucleotides.

The invention further provides a method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide selected from the group consisting of a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:13-24, b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:13-24, c) a polynucleotide complementary to the polynucleotide of a), d) a polynucleotide complementary to the polynucleotide of b), and e) an RNA equivalent of a)-d). The method comprises a) amplifying said target polynucleotide or fragment thereof using polymerase chain reaction amplification, and b) detecting the presence or absence of said amplified target polynucleotide or fragment thereof, and, optionally, if present, the amount thereof.

The invention further provides a composition comprising an effective amount of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-12, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-12, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-12, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-12, and a pharmaceutically acceptable excipient. In one embodiment, the composition comprises an amino acid sequence selected from the group consisting of SEQ ID NO:1-12. The invention additionally provides a method of treating a disease or condition associated with decreased expression of functional IGSFP, comprising administering to a patient in need of such treatment the composition.

The invention also provides a method for screening a compound for effectiveness as an agonist of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-12, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-12, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-12, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-12. The method comprises a) exposing a sample comprising the polypeptide to a compound, and b) detecting agonist activity in the sample. In one alternative, the invention provides a composition comprising an agonist compound identified by the method and a pharmaceutically acceptable excipient. In another alternative, the invention provides a method of treating a disease or condition associated with decreased expression of functional IGSFP, comprising

administering to a patient in need of such treatment the composition.

Additionally, the invention provides a method for screening a compound for effectiveness as an antagonist of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-12, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-12, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-12, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-12. The method comprises a) exposing a sample comprising the polypeptide to a compound, and b) detecting antagonist activity in the sample. In one alternative, the invention provides a composition comprising an antagonist compound identified by the method and a pharmaceutically acceptable excipient. In another alternative, the invention provides a method of treating a disease or condition associated with overexpression of functional IGSFP, comprising administering to a patient in need of such treatment the composition.

The invention further provides a method of screening for a compound that specifically binds to a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-12, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-12, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-12, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-12. The method comprises a) combining the polypeptide with at least one test compound under suitable conditions, and b) detecting binding of the polypeptide to the test compound, thereby identifying a compound that specifically binds to the polypeptide.

The invention further provides a method of screening for a compound that modulates the activity of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-12, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-12, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-12, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-12. The method comprises a) combining the polypeptide with at least one test compound under conditions permissive for the activity of the polypeptide, b) assessing the activity

of the polypeptide in the presence of the test compound, and c) comparing the activity of the polypeptide in the presence of the test compound with the activity of the polypeptide in the absence of the test compound, wherein a change in the activity of the polypeptide in the presence of the test compound is indicative of a compound that modulates the activity of the polypeptide.

The invention further provides a method for screening a compound for effectiveness in altering expression of a target polynucleotide, wherein said target polynucleotide comprises a polynucleotide sequence selected from the group consisting of SEQ ID NO:13-24, the method comprising a) exposing a sample comprising the target polynucleotide to a compound, b) detecting altered expression of the target polynucleotide, and c) comparing the expression of the target polynucleotide in the presence of varying amounts of the compound and in the absence of the compound.

The invention further provides a method for assessing toxicity of a test compound, said method comprising a) treating a biological sample containing nucleic acids with the test compound; b) hybridizing the nucleic acids of the treated biological sample with a probe comprising at least 20 contiguous nucleotides of a polynucleotide selected from the group consisting of i) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:13-24, ii) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:13-24, iii) a polynucleotide having a sequence complementary to i), iv) a polynucleotide complementary to the polynucleotide of ii), and v) an RNA equivalent of i)-iv). Hybridization occurs under conditions whereby a specific hybridization complex is formed between said probe and a target polynucleotide in the biological sample, said target polynucleotide selected from the group consisting of i) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:13-24, ii) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:13-24, iii) a polynucleotide complementary to the polynucleotide of i), iv) a polynucleotide complementary to the polynucleotide of ii), and v) an RNA equivalent of i)-iv). Alternatively, the target polynucleotide comprises a fragment of a polynucleotide sequence selected from the group consisting of i)-v) above; c) quantifying the amount of hybridization complex; and d) comparing the amount of hybridization complex in the treated biological sample with the amount of hybridization complex in an untreated biological sample, wherein a difference in the amount of hybridization complex in the treated biological sample is indicative of toxicity of the test compound.

BRIEF DESCRIPTION OF THE TABLES

Table 1 summarizes the nomenclature for the full length polynucleotide and polypeptide sequences of the present invention.

Table 2 shows the GenBank identification number and annotation of the nearest GenBank homolog, and the PROTEOME database identification numbers and annotations of PROTEOME database homologs, for polypeptides of the invention. The probability scores for the matches between each polypeptide and its homolog(s) are also shown.

Table 3 shows structural features of polypeptide sequences of the invention, including predicted motifs and domains, along with the methods, algorithms, and searchable databases used for analysis of the polypeptides.

Table 4 lists the cDNA and/or genomic DNA fragments which were used to assemble polynucleotide sequences of the invention, along with selected fragments of the polynucleotide sequences.

Table 5 shows the representative cDNA library for polynucleotides of the invention.

Table 6 provides an appendix which describes the tissues and vectors used for construction of the cDNA libraries shown in Table 5.

Table 7 shows the tools, programs, and algorithms used to analyze the polynucleotides and polypeptides of the invention, along with applicable descriptions, references, and threshold parameters.

DESCRIPTION OF THE INVENTION

Before the present proteins, nucleotide sequences, and methods are described, it is understood that this invention is not limited to the particular machines, materials and methods described, as these may vary. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments only, and is not intended to limit the scope of the present invention which will be limited only by the appended claims.

It must be noted that as used herein and in the appended claims, the singular forms "a," "an," and "the" include plural reference unless the context clearly dictates otherwise. Thus, for example, a reference to "a host cell" includes a plurality of such host cells, and a reference to "an antibody" is a reference to one or more antibodies and equivalents thereof known to those skilled in the art, and so forth.

Unless defined otherwise, all technical and scientific terms used herein have the same meanings as commonly understood by one of ordinary skill in the art to which this invention belongs. Although any machines, materials, and methods similar or equivalent to those described herein can be

used to practice or test the present invention, the preferred machines, materials and methods are now described. All publications mentioned herein are cited for the purpose of describing and disclosing the cell lines, protocols, reagents and vectors which are reported in the publications and which might be used in connection with the invention. Nothing herein is to be construed as an admission that the invention is not entitled to antedate such disclosure by virtue of prior invention.

DEFINITIONS

"IGSFP" refers to the amino acid sequences of substantially purified IGSFP obtained from any species, particularly a mammalian species, including bovine, ovine, porcine, murine, equine, and human, and from any source, whether natural, synthetic, semi-synthetic, or recombinant.

The term "agonist" refers to a molecule which intensifies or mimics the biological activity of IGSFP. Agonists may include proteins, nucleic acids, carbohydrates, small molecules, or any other compound or composition which modulates the activity of IGSFP either by directly interacting with IGSFP or by acting on components of the biological pathway in which IGSFP participates.

An "allelic variant" is an alternative form of the gene encoding IGSFP. Allelic variants may result from at least one mutation in the nucleic acid sequence and may result in altered mRNAs or in polypeptides whose structure or function may or may not be altered. A gene may have none, one, or many allelic variants of its naturally occurring form. Common mutational changes which give rise to allelic variants are generally ascribed to natural deletions, additions, or substitutions of nucleotides. Each of these types of changes may occur alone, or in combination with the others, one or more times in a given sequence.

"Altered" nucleic acid sequences encoding IGSFP include those sequences with deletions, insertions, or substitutions of different nucleotides, resulting in a polypeptide the same as IGSFP or a polypeptide with at least one functional characteristic of IGSFP. Included within this definition are polymorphisms which may or may not be readily detectable using a particular oligonucleotide probe of the polynucleotide encoding IGSFP, and improper or unexpected hybridization to allelic variants, with a locus other than the normal chromosomal locus for the polynucleotide sequence encoding IGSFP. The encoded protein may also be "altered," and may contain deletions, insertions, or substitutions of amino acid residues which produce a silent change and result in a functionally equivalent IGSFP. Deliberate amino acid substitutions may be made on the basis of similarity in polarity, charge, solubility, hydrophobicity, hydrophilicity, and/or the amphipathic nature of the residues, as long as the biological or immunological activity of IGSFP is retained. For example, negatively charged amino acids may include aspartic acid and glutamic acid, and positively charged amino acids may include lysine and arginine. Amino acids with uncharged polar side chains having similar hydrophilicity values may

include: asparagine and glutamine; and serine and threonine. Amino acids with uncharged side chains having similar hydrophilicity values may include: leucine, isoleucine, and valine; glycine and alanine; and phenylalanine and tyrosine.

The terms "amino acid" and "amino acid sequence" refer to an oligopeptide, peptide, polypeptide, or protein sequence, or a fragment of any of these, and to naturally occurring or synthetic molecules. Where "amino acid sequence" is recited to refer to a sequence of a naturally occurring protein molecule, "amino acid sequence" and like terms are not meant to limit the amino acid sequence to the complete native amino acid sequence associated with the recited protein molecule.

"Amplification" relates to the production of additional copies of a nucleic acid sequence. Amplification is generally carried out using polymerase chain reaction (PCR) technologies well known in the art.

The term "antagonist" refers to a molecule which inhibits or attenuates the biological activity of IGSFP. Antagonists may include proteins such as antibodies, nucleic acids, carbohydrates, small molecules, or any other compound or composition which modulates the activity of IGSFP either by directly interacting with IGSFP or by acting on components of the biological pathway in which IGSFP participates.

The term "antibody" refers to intact immunoglobulin molecules as well as to fragments thereof, such as Fab, $F(ab')_2$, and Fv fragments, which are capable of binding an epitopic determinant. Antibodies that bind IGSFP polypeptides can be prepared using intact polypeptides or using fragments containing small peptides of interest as the immunizing antigen. The polypeptide or oligopeptide used to immunize an animal (e.g., a mouse, a rat, or a rabbit) can be derived from the translation of RNA, or synthesized chemically, and can be conjugated to a carrier protein if desired. Commonly used carriers that are chemically coupled to peptides include bovine serum albumin, thyroglobulin, and keyhole limpet hemocyanin (KLH). The coupled peptide is then used to immunize the animal.

The term "antigenic determinant" refers to that region of a molecule (i.e., an epitope) that makes contact with a particular antibody. When a protein or a fragment of a protein is used to immunize a host animal, numerous regions of the protein may induce the production of antibodies which bind specifically to antigenic determinants (particular regions or three-dimensional structures on the protein). An antigenic determinant may compete with the intact antigen (i.e., the immunogen used to elicit the immune response) for binding to an antibody.

The term "aptamer" refers to a nucleic acid or oligonucleotide molecule that binds to a specific molecular target. Aptamers are derived from an in vitro evolutionary process (e.g., SELEX (Systematic Evolution of Ligands by EXponential Enrichment), described in U.S. Patent No.

5,270,163), which selects for target-specific aptamer sequences from large combinatorial libraries. Aptamer compositions may be double-stranded or single-stranded, and may include deoxyribonucleotides, ribonucleotides, nucleotide derivatives, or other nucleotide-like molecules. The nucleotide components of an aptamer may have modified sugar groups (e.g., the 2'-OH group of a ribonucleotide may be replaced by 2'-F or 2'-NH_2), which may improve a desired property, e.g., resistance to nucleases or longer lifetime in blood. Aptamers may be conjugated to other molecules, e.g., a high molecular weight carrier to slow clearance of the aptamer from the circulatory system. Aptamers may be specifically cross-linked to their cognate ligands, e.g., by photo-activation of a cross-linker. (See, e.g., Brody, E.N. and L. Gold (2000) J. Biotechnol. 74:5-13.)

The term "intramer" refers to an aptamer which is expressed in vivo. For example, a vaccinia virus-based RNA expression system has been used to express specific RNA aptamers at high levels in the cytoplasm of leukocytes (Blind, M. et al. (1999) Proc. Natl Acad. Sci. USA 96:3606-3610).

The term "spiegelmer" refers to an aptamer which includes L-DNA, L-RNA, or other left-handed nucleotide derivatives or nucleotide-like molecules. Aptamers containing left-handed nucleotides are resistant to degradation by naturally occurring enzymes, which normally act on substrates containing right-handed nucleotides.

The term "antisense" refers to any composition capable of base-pairing with the "sense" (coding) strand of a specific nucleic acid sequence. Antisense compositions may include DNA; RNA; peptide nucleic acid (PNA); oligonucleotides having modified backbone linkages such as phosphorothioates, methylphosphonates, or benzylphosphonates; oligonucleotides having modified sugar groups such as 2'-methoxyethyl sugars or 2'-methoxyethoxy sugars; or oligonucleotides having modified bases such as 5-methyl cytosine, 2'-deoxyuracil, or 7-deaza-2'-deoxyguanosine. Antisense molecules may be produced by any method including chemical synthesis or transcription. Once introduced into a cell, the complementary antisense molecule base-pairs with a naturally occurring nucleic acid sequence produced by the cell to form duplexes which block either transcription or translation. The designation "negative" or "minus" can refer to the antisense strand, and the designation "positive" or "plus" can refer to the sense strand of a reference DNA molecule.

The term "biologically active" refers to a protein having structural, regulatory, or biochemical functions of a naturally occurring molecule. Likewise, "immunologically active" or "immunogenic" refers to the capability of the natural, recombinant, or synthetic IGSFP, or of any oligopeptide thereof, to induce a specific immune response in appropriate animals or cells and to bind with specific antibodies.

"Complementary" describes the relationship between two single-stranded nucleic acid

sequences that anneal by base-pairing. For example, 5'-AGT-3' pairs with its complement, 3'-TCA-5'.

A "composition comprising a given polynucleotide sequence" and a "composition comprising a given amino acid sequence" refer broadly to any composition containing the given polynucleotide or amino acid sequence. The composition may comprise a dry formulation or an aqueous solution. Compositions comprising polynucleotide sequences encoding IGSFP or fragments of IGSFP may be employed as hybridization probes. The probes may be stored in freeze-dried form and may be associated with a stabilizing agent such as a carbohydrate. In hybridizations, the probe may be deployed in an aqueous solution containing salts (e.g., NaCl), detergents (e.g., sodium dodecyl sulfate; SDS), and other components (e.g., Denhardt's solution, dry milk, salmon sperm DNA, etc.).

"Consensus sequence" refers to a nucleic acid sequence which has been subjected to repeated DNA sequence analysis to resolve uncalled bases, extended using the XL-PCR kit (Applied Biosystems, Foster City CA) in the 5' and/or the 3' direction, and resequenced, or which has been assembled from one or more overlapping cDNA, EST, or genomic DNA fragments using a computer program for fragment assembly, such as the GELVIEW fragment assembly system (GCG, Madison WI) or Phrap (University of Washington, Seattle WA). Some sequences have been both extended and assembled to produce the consensus sequence.

"Conservative amino acid substitutions" are those substitutions that are predicted to least interfere with the properties of the original protein, i.e., the structure and especially the function of the protein is conserved and not significantly changed by such substitutions. The table below shows amino acids which may be substituted for an original amino acid in a protein and which are regarded as conservative amino acid substitutions.

Original Residue	Conservative Substitution
Ala	Gly, Ser
Arg	His, Lys
Asn	Asp, Gln, His
Asp	Asn, Glu
Cys	Ala, Ser
Gln	Asn, Glu, His
Glu	Asp, Gln, His
Gly	Ala
His	Asn, Arg, Gln, Glu
Ile	Leu, Val
Leu	Ile, Val
Lys	Arg, Gln, Glu
Met	Leu, Ile
Phe	His, Met, Leu, Trp, Tyr
Ser	Cys, Thr

Thr	Ser, Val
Trp	Phe, Tyr
Tyr	His, Phe, Trp
Val	Ile, Leu, Thr

Conservative amino acid substitutions generally maintain (a) the structure of the polypeptide backbone in the area of the substitution, for example, as a beta sheet or alpha helical conformation, (b) the charge or hydrophobicity of the molecule at the site of the substitution, and/or (c) the bulk of the side chain.

A "deletion" refers to a change in the amino acid or nucleotide sequence that results in the absence of one or more amino acid residues or nucleotides.

The term "derivative" refers to a chemically modified polynucleotide or polypeptide. Chemical modifications of a polynucleotide can include, for example, replacement of hydrogen by an alkyl, acyl, hydroxyl, or amino group. A derivative polynucleotide encodes a polypeptide which retains at least one biological or immunological function of the natural molecule. A derivative polypeptide is one modified by glycosylation, pegylation, or any similar process that retains at least one biological or immunological function of the polypeptide from which it was derived.

A "detectable label" refers to a reporter molecule or enzyme that is capable of generating a measurable signal and is covalently or noncovalently joined to a polynucleotide or polypeptide.

"Differential expression" refers to increased or upregulated; or decreased, downregulated, or absent gene or protein expression, determined by comparing at least two different samples. Such comparisons may be carried out between, for example, a treated and an untreated sample, or a diseased and a normal sample.

"Exon shuffling" refers to the recombination of different coding regions (exons). Since an exon may represent a structural or functional domain of the encoded protein, new proteins may be assembled through the novel reassortment of stable substructures, thus allowing acceleration of the evolution of new protein functions.

A "fragment" is a unique portion of IGSFP or the polynucleotide encoding IGSFP which is identical in sequence to but shorter in length than the parent sequence. A fragment may comprise up to the entire length of the defined sequence, minus one nucleotide/amino acid residue. For example, a fragment may comprise from 5 to 1000 contiguous nucleotides or amino acid residues. A fragment used as a probe, primer, antigen, therapeutic molecule, or for other purposes, may be at least 5, 10, 15, 16, 20, 25, 30, 40, 50, 60, 75, 100, 150, 250 or at least 500 contiguous nucleotides or amino acid residues in length. Fragments may be preferentially selected from certain regions of a molecule. For example, a polypeptide fragment may comprise a certain length of contiguous amino acids selected

from the first 250 or 500 amino acids (or first 25% or 50%) of a polypeptide as shown in a certain defined sequence. Clearly these lengths are exemplary, and any length that is supported by the specification, including the Sequence Listing, tables, and figures, may be encompassed by the present embodiments.

A fragment of SEQ ID NO:13-24 comprises a region of unique polynucleotide sequence that specifically identifies SEQ ID NO:13-24, for example, as distinct from any other sequence in the genome from which the fragment was obtained. A fragment of SEQ ID NO:13-24 is useful, for example, in hybridization and amplification technologies and in analogous methods that distinguish SEQ ID NO:13-24 from related polynucleotide sequences. The precise length of a fragment of SEQ ID NO:13-24 and the region of SEQ ID NO:13-24 to which the fragment corresponds are routinely determinable by one of ordinary skill in the art based on the intended purpose for the fragment.

A fragment of SEQ ID NO:1-12 is encoded by a fragment of SEQ ID NO:13-24. A fragment of SEQ ID NO:1-12 comprises a region of unique amino acid sequence that specifically identifies SEQ ID NO:1-12. For example, a fragment of SEQ ID NO:1-12 is useful as an immunogenic peptide for the development of antibodies that specifically recognize SEQ ID NO:1-12. The precise length of a fragment of SEQ ID NO:1-12 and the region of SEQ ID NO:1-12 to which the fragment corresponds are routinely determinable by one of ordinary skill in the art based on the intended purpose for the fragment.

A "full length" polynucleotide sequence is one containing at least a translation initiation codon (e.g., methionine) followed by an open reading frame and a translation termination codon. A "full length" polynucleotide sequence encodes a "full length" polypeptide sequence.

"Homology" refers to sequence similarity or, interchangeably, sequence identity, between two or more polynucleotide sequences or two or more polypeptide sequences.

The terms "percent identity" and "% identity," as applied to polynucleotide sequences, refer to the percentage of residue matches between at least two polynucleotide sequences aligned using a standardized algorithm. Such an algorithm may insert, in a standardized and reproducible way, gaps in the sequences being compared in order to optimize alignment between two sequences, and therefore achieve a more meaningful comparison of the two sequences.

Percent identity between polynucleotide sequences may be determined using the default parameters of the CLUSTAL V algorithm as incorporated into the MEGALIGN version 3.12e sequence alignment program. This program is part of the LASERGENE software package, a suite of molecular biological analysis programs (DNASTAR, Madison WI). CLUSTAL V is described in Higgins, D.G. and P.M. Sharp (1989) CABIOS 5:151-153 and in Higgins, D.G. et al. (1992) CABIOS

8:189-191. For pairwise alignments of polynucleotide sequences, the default parameters are set as follows: Ktuple=2, gap penalty=5, window=4, and "diagonals saved"=4. The "weighted" residue weight table is selected as the default. Percent identity is reported by CLUSTAL V as the "percent similarity" between aligned polynucleotide sequences.

Alternatively, a suite of commonly used and freely available sequence comparison algorithms is provided by the National Center for Biotechnology Information (NCBI) Basic Local Alignment Search Tool (BLAST) (Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410), which is available from several sources, including the NCBI, Bethesda, MD, and on the Internet at http://www.ncbi.nlm.nih.gov/BLAST/. The BLAST software suite includes various sequence analysis programs including "blastn," that is used to align a known polynucleotide sequence with other polynucleotide sequences from a variety of databases. Also available is a tool called "BLAST 2 Sequences" that is used for direct pairwise comparison of two nucleotide sequences. "BLAST 2 Sequences" can be accessed and used interactively at http://www.ncbi.nlm.nih.gov/gorf/bl2.html. The "BLAST 2 Sequences" tool can be used for both blastn and blastp (discussed below). BLAST programs are commonly used with gap and other parameters set to default settings. For example, to compare two nucleotide sequences, one may use blastn with the "BLAST 2 Sequences" tool Version 2.0.12 (April-21-2000) set at default parameters. Such default parameters may be, for example:

Matrix: BLOSUM62

Reward for match: 1

Penalty for mismatch: -2

Open Gap: 5 and Extension Gap: 2 penalties

Gap x drop-off: 50

Expect: 10

Word Size: 11

Filter: on

Percent identity may be measured over the length of an entire defined sequence, for example, as defined by a particular SEQ ID number, or may be measured over a shorter length, for example, over the length of a fragment taken from a larger, defined sequence, for instance, a fragment of at least 20, at least 30, at least 40, at least 50, at least 70, at least 100, or at least 200 contiguous nucleotides. Such lengths are exemplary only, and it is understood that any fragment length supported by the sequences shown herein, in the tables, figures, or Sequence Listing, may be used to describe a length over which percentage identity may be measured.

Nucleic acid sequences that do not show a high degree of identity may nevertheless encode

similar amino acid sequences due to the degeneracy of the genetic code. It is understood that changes in a nucleic acid sequence can be made using this degeneracy to produce multiple nucleic acid sequences that all encode substantially the same protein.

The phrases "percent identity" and "% identity," as applied to polypeptide sequences, refer to the percentage of residue matches between at least two polypeptide sequences aligned using a standardized algorithm. Methods of polypeptide sequence alignment are well-known. Some alignment methods take into account conservative amino acid substitutions. Such conservative substitutions, explained in more detail above, generally preserve the charge and hydrophobicity at the site of substitution, thus preserving the structure (and therefore function) of the polypeptide.

Percent identity between polypeptide sequences may be determined using the default parameters of the CLUSTAL V algorithm as incorporated into the MEGALIGN version 3.12e sequence alignment program (described and referenced above). For pairwise alignments of polypeptide sequences using CLUSTAL V, the default parameters are set as follows: Ktuple=1, gap penalty=3, window=5, and "diagonals saved"=5. The PAM250 matrix is selected as the default residue weight table. As with polynucleotide alignments, the percent identity is reported by CLUSTAL V as the "percent similarity" between aligned polypeptide sequence pairs.

Alternatively the NCBI BLAST software suite may be used. For example, for a pairwise comparison of two polypeptide sequences, one may use the "BLAST 2 Sequences" tool Version 2.0.12 (April-21-2000) with blastp set at default parameters. Such default parameters may be, for example:

Matrix: BLOSUM62
Open Gap: 11 and Extension Gap: 1 penalties
Gap x drop-off: 50
Expect: 10
Word Size: 3
Filter: on

Percent identity may be measured over the length of an entire defined polypeptide sequence, for example, as defined by a particular SEQ ID number, or may be measured over a shorter length, for example, over the length of a fragment taken from a larger, defined polypeptide sequence, for instance, a fragment of at least 15, at least 20, at least 30, at least 40, at least 50, at least 70 or at least 150 contiguous residues. Such lengths are exemplary only, and it is understood that any fragment length supported by the sequences shown herein, in the tables, figures or Sequence Listing, may be used to describe a length over which percentage identity may be measured.

"Human artificial chromosomes" (HACs) are linear microchromosomes which may contain DNA sequences of about 6 kb to 10 Mb in size and which contain all of the elements required for chromosome replication, segregation and maintenance.

The term "humanized antibody" refers to an antibody molecule in which the amino acid sequence in the non-antigen binding regions has been altered so that the antibody more closely resembles a human antibody, and still retains its original binding ability.

"Hybridization" refers to the process by which a polynucleotide strand anneals with a complementary strand through base pairing under defined hybridization conditions. Specific hybridization is an indication that two nucleic acid sequences share a high degree of complementarity. Specific hybridization complexes form under permissive annealing conditions and remain hybridized after the "washing" step(s). The washing step(s) is particularly important in determining the stringency of the hybridization process, with more stringent conditions allowing less non-specific binding, i.e., binding between pairs of nucleic acid strands that are not perfectly matched. Permissive conditions for annealing of nucleic acid sequences are routinely determinable by one of ordinary skill in the art and may be consistent among hybridization experiments, whereas wash conditions may be varied among experiments to achieve the desired stringency, and therefore hybridization specificity. Permissive annealing conditions occur, for example, at 68°C in the presence of about 6 x SSC, about 1% (w/v) SDS, and about 100 μg/ml sheared, denatured salmon sperm DNA.

Generally, stringency of hybridization is expressed, in part, with reference to the temperature under which the wash step is carried out. Such wash temperatures are typically selected to be about 5°C to 20°C lower than the thermal melting point (T_m) for the specific sequence at a defined ionic strength and pH. The T_m is the temperature (under defined ionic strength and pH) at which 50% of the target sequence hybridizes to a perfectly matched probe. An equation for calculating T_m and conditions for nucleic acid hybridization are well known and can be found in Sambrook, J. et al. (1989) Molecular Cloning: A Laboratory Manual, 2nd ed., vol. 1-3, Cold Spring Harbor Press, Plainview NY; specifically see volume 2, chapter 9.

High stringency conditions for hybridization between polynucleotides of the present invention include wash conditions of 68°C in the presence of about 0.2 x SSC and about 0.1% SDS, for 1 hour. Alternatively, temperatures of about 65°C, 60°C, 55°C, or 42°C may be used. SSC concentration may be varied from about 0.1 to 2 x SSC, with SDS being present at about 0.1%. Typically, blocking reagents are used to block non-specific hybridization. Such blocking reagents include, for instance, sheared and denatured salmon sperm DNA at about 100-200 μg/ml. Organic solvent, such as formamide at a concentration of about 35-50% v/v, may also be used under particular circumstances,

such as for RNA:DNA hybridizations. Useful variations on these wash conditions will be readily apparent to those of ordinary skill in the art. Hybridization, particularly under high stringency conditions, may be suggestive of evolutionary similarity between the nucleotides. Such similarity is strongly indicative of a similar role for the nucleotides and their encoded polypeptides.

The term "hybridization complex" refers to a complex formed between two nucleic acid sequences by virtue of the formation of hydrogen bonds between complementary bases. A hybridization complex may be formed in solution (e.g., C_0t or R_0t analysis) or formed between one nucleic acid sequence present in solution and another nucleic acid sequence immobilized on a solid support (e.g., paper, membranes, filters, chips, pins or glass slides, or any other appropriate substrate to which cells or their nucleic acids have been fixed).

The words "insertion" and "addition" refer to changes in an amino acid or nucleotide sequence resulting in the addition of one or more amino acid residues or nucleotides, respectively.

"Immune response" can refer to conditions associated with inflammation, trauma, immune disorders, or infectious or genetic disease, etc. These conditions can be characterized by expression of various factors, e.g., cytokines, chemokines, and other signaling molecules, which may affect cellular and systemic defense systems.

An "immunogenic fragment" is a polypeptide or oligopeptide fragment of IGSFP which is capable of eliciting an immune response when introduced into a living organism, for example, a mammal. The term "immunogenic fragment" also includes any polypeptide or oligopeptide fragment of IGSFP which is useful in any of the antibody production methods disclosed herein or known in the art.

The term "microarray" refers to an arrangement of a plurality of polynucleotides, polypeptides, or other chemical compounds on a substrate.

The terms "element" and "array element" refer to a polynucleotide, polypeptide, or other chemical compound having a unique and defined position on a microarray.

The term "modulate" refers to a change in the activity of IGSFP. For example, modulation may cause an increase or a decrease in protein activity, binding characteristics, or any other biological, functional, or immunological properties of IGSFP.

The phrases "nucleic acid" and "nucleic acid sequence" refer to a nucleotide, oligonucleotide, polynucleotide, or any fragment thereof. These phrases also refer to DNA or RNA of genomic or synthetic origin which may be single-stranded or double-stranded and may represent the sense or the antisense strand, to peptide nucleic acid (PNA), or to any DNA-like or RNA-like material.

"Operably linked" refers to the situation in which a first nucleic acid sequence is placed in a

functional relationship with a second nucleic acid sequence. For instance, a promoter is operably linked to a coding sequence if the promoter affects the transcription or expression of the coding sequence. Operably linked DNA sequences may be in close proximity or contiguous and, where necessary to join two protein coding regions, in the same reading frame.

"Peptide nucleic acid" (PNA) refers to an antisense molecule or anti-gene agent which comprises an oligonucleotide of at least about 5 nucleotides in length linked to a peptide backbone of amino acid residues ending in lysine. The terminal lysine confers solubility to the composition. PNAs preferentially bind complementary single stranded DNA or RNA and stop transcript elongation, and may be pegylated to extend their lifespan in the cell.

"Post-translational modification" of an IGSFP may involve lipidation, glycosylation, phosphorylation, acetylation, racemization, proteolytic cleavage, and other modifications known in the art. These processes may occur synthetically or biochemically. Biochemical modifications will vary by cell type depending on the enzymatic milieu of IGSFP.

"Probe" refers to nucleic acid sequences encoding IGSFP, their complements, or fragments thereof, which are used to detect identical, allelic or related nucleic acid sequences. Probes are isolated oligonucleotides or polynucleotides attached to a detectable label or reporter molecule. Typical labels include radioactive isotopes, ligands, chemiluminescent agents, and enzymes. "Primers" are short nucleic acids, usually DNA oligonucleotides, which may be annealed to a target polynucleotide by complementary base-pairing. The primer may then be extended along the target DNA strand by a DNA polymerase enzyme. Primer pairs can be used for amplification (and identification) of a nucleic acid sequence, e.g., by the polymerase chain reaction (PCR).

Probes and primers as used in the present invention typically comprise at least 15 contiguous nucleotides of a known sequence. In order to enhance specificity, longer probes and primers may also be employed, such as probes and primers that comprise at least 20, 25, 30, 40, 50, 60, 70, 80, 90, 100, or at least 150 consecutive nucleotides of the disclosed nucleic acid sequences. Probes and primers may be considerably longer than these examples, and it is understood that any length supported by the specification, including the tables, figures, and Sequence Listing, may be used.

Methods for preparing and using probes and primers are described in the references, for example Sambrook, J. et al. (1989) Molecular Cloning: A Laboratory Manual, 2nd ed., vol. 1-3, Cold Spring Harbor Press, Plainview NY; Ausubel, F.M. et al. (1987) Current Protocols in Molecular Biology, Greene Publ. Assoc. & Wiley-Intersciences, New York NY; Innis, M. et al. (1990) PCR Protocols, A Guide to Methods and Applications, Academic Press, San Diego CA. PCR primer pairs can be derived from a known sequence, for example, by using computer programs intended for that

purpose such as Primer (Version 0.5, 1991, Whitehead Institute for Biomedical Research, Cambridge MA).

Oligonucleotides for use as primers are selected using software known in the art for such purpose. For example, OLIGO 4.06 software is useful for the selection of PCR primer pairs of up to 100 nucleotides each, and for the analysis of oligonucleotides and larger polynucleotides of up to 5,000 nucleotides from an input polynucleotide sequence of up to 32 kilobases. Similar primer selection programs have incorporated additional features for expanded capabilities. For example, the PrimOU primer selection program (available to the public from the Genome Center at University of Texas South West Medical Center, Dallas TX) is capable of choosing specific primers from megabase sequences and is thus useful for designing primers on a genome-wide scope. The Primer3 primer selection program (available to the public from the Whitehead Institute/MIT Center for Genome Research, Cambridge MA) allows the user to input a "mispriming library," in which sequences to avoid as primer binding sites are user-specified. Primer3 is useful, in particular, for the selection of oligonucleotides for microarrays. (The source code for the latter two primer selection programs may also be obtained from their respective sources and modified to meet the user's specific needs.) The PrimeGen program (available to the public from the UK Human Genome Mapping Project Resource Centre, Cambridge UK) designs primers based on multiple sequence alignments, thereby allowing selection of primers that hybridize to either the most conserved or least conserved regions of aligned nucleic acid sequences. Hence, this program is useful for identification of both unique and conserved oligonucleotides and polynucleotide fragments. The oligonucleotides and polynucleotide fragments identified by any of the above selection methods are useful in hybridization technologies, for example, as PCR or sequencing primers, microarray elements, or specific probes to identify fully or partially complementary polynucleotides in a sample of nucleic acids. Methods of oligonucleotide selection are not limited to those described above.

A "recombinant nucleic acid" is a sequence that is not naturally occurring or has a sequence that is made by an artificial combination of two or more otherwise separated segments of sequence. This artificial combination is often accomplished by chemical synthesis or, more commonly, by the artificial manipulation of isolated segments of nucleic acids, e.g., by genetic engineering techniques such as those described in Sambrook, supra. The term recombinant includes nucleic acids that have been altered solely by addition, substitution, or deletion of a portion of the nucleic acid. Frequently, a recombinant nucleic acid may include a nucleic acid sequence operably linked to a promoter sequence. Such a recombinant nucleic acid may be part of a vector that is used, for example, to transform a cell.

Alternatively, such recombinant nucleic acids may be part of a viral vector, e.g., based on a

vaccinia virus, that could be use to vaccinate a mammal wherein the recombinant nucleic acid is expressed, inducing a protective immunological response in the mammal.

A "regulatory element" refers to a nucleic acid sequence usually derived from untranslated regions of a gene and includes enhancers, promoters, introns, and 5' and 3' untranslated regions (UTRs). Regulatory elements interact with host or viral proteins which control transcription, translation, or RNA stability.

"Reporter molecules" are chemical or biochemical moieties used for labeling a nucleic acid, amino acid, or antibody. Reporter molecules include radionuclides; enzymes; fluorescent, chemiluminescent, or chromogenic agents; substrates; cofactors; inhibitors; magnetic particles; and other moieties known in the art.

An "RNA equivalent," in reference to a DNA sequence, is composed of the same linear sequence of nucleotides as the reference DNA sequence with the exception that all occurrences of the nitrogenous base thymine are replaced with uracil, and the sugar backbone is composed of ribose instead of deoxyribose.

The term "sample" is used in its broadest sense. A sample suspected of containing IGSFP, nucleic acids encoding IGSFP, or fragments thereof may comprise a bodily fluid; an extract from a cell, chromosome, organelle, or membrane isolated from a cell; a cell; genomic DNA, RNA, or cDNA, in solution or bound to a substrate; a tissue; a tissue print; etc.

The terms "specific binding" and "specifically binding" refer to that interaction between a protein or peptide and an agonist, an antibody, an antagonist, a small molecule, or any natural or synthetic binding composition. The interaction is dependent upon the presence of a particular structure of the protein, e.g., the antigenic determinant or epitope, recognized by the binding molecule. For example, if an antibody is specific for epitope "A," the presence of a polypeptide comprising the epitope A, or the presence of free unlabeled A, in a reaction containing free labeled A and the antibody will reduce the amount of labeled A that binds to the antibody.

The term "substantially purified" refers to nucleic acid or amino acid sequences that are removed from their natural environment and are isolated or separated, and are at least 60% free, preferably at least 75% free, and most preferably at least 90% free from other components with which they are naturally associated.

A "substitution" refers to the replacement of one or more amino acid residues or nucleotides by different amino acid residues or nucleotides, respectively.

"Substrate" refers to any suitable rigid or semi-rigid support including membranes, filters, chips, slides, wafers, fibers, magnetic or nonmagnetic beads, gels, tubing, plates, polymers,

microparticles and capillaries. The substrate can have a variety of surface forms, such as wells, trenches, pins, channels and pores, to which polynucleotides or polypeptides are bound.

A "transcript image" or "expression profile" refers to the collective pattern of gene expression by a particular cell type or tissue under given conditions at a given time.

"Transformation" describes a process by which exogenous DNA is introduced into a recipient cell. Transformation may occur under natural or artificial conditions according to various methods well known in the art, and may rely on any known method for the insertion of foreign nucleic acid sequences into a prokaryotic or eukaryotic host cell. The method for transformation is selected based on the type of host cell being transformed and may include, but is not limited to, bacteriophage or viral infection, electroporation, heat shock, lipofection, and particle bombardment. The term "transformed cells" includes stably transformed cells in which the inserted DNA is capable of replication either as an autonomously replicating plasmid or as part of the host chromosome, as well as transiently transformed cells which express the inserted DNA or RNA for limited periods of time.

A "transgenic organism," as used herein, is any organism, including but not limited to animals and plants, in which one or more of the cells of the organism contains heterologous nucleic acid introduced by way of human intervention, such as by transgenic techniques well known in the art. The nucleic acid is introduced into the cell, directly or indirectly by introduction into a precursor of the cell, by way of deliberate genetic manipulation, such as by microinjection or by infection with a recombinant virus. In one alternative, the nucleic acid can be introduced by infection with a recombinant viral vector, such as a lentiviral vector (Lois, C. et al. (2002) Science 295:868-872). The term genetic manipulation does not include classical cross-breeding, or in vitro fertilization, but rather is directed to the introduction of a recombinant DNA molecule. The transgenic organisms contemplated in accordance with the present invention include bacteria, cyanobacteria, fungi, plants and animals. The isolated DNA of the present invention can be introduced into the host by methods known in the art, for example infection, transfection, transformation or transconjugation. Techniques for transferring the DNA of the present invention into such organisms are widely known and provided in references such as Sambrook et al. (1989), supra.

A "variant" of a particular nucleic acid sequence is defined as a nucleic acid sequence having at least 40% sequence identity to the particular nucleic acid sequence over a certain length of one of the nucleic acid sequences using blastn with the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-1999) set at default parameters. Such a pair of nucleic acids may show, for example, at least 50%, at least 60%, at least 70%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% or greater

sequence identity over a certain defined length. A variant may be described as, for example, an "allelic" (as defined above), "splice," "species," or "polymorphic" variant. A splice variant may have significant identity to a reference molecule, but will generally have a greater or lesser number of polynucleotides due to alternate splicing of exons during mRNA processing. The corresponding polypeptide may possess additional functional domains or lack domains that are present in the reference molecule. Species variants are polynucleotide sequences that vary from one species to another. The resulting polypeptides will generally have significant amino acid identity relative to each other. A polymorphic variant is a variation in the polynucleotide sequence of a particular gene between individuals of a given species. Polymorphic variants also may encompass "single nucleotide polymorphisms" (SNPs) in which the polynucleotide sequence varies by one nucleotide base. The presence of SNPs may be indicative of, for example, a certain population, a disease state, or a propensity for a disease state.

A "variant" of a particular polypeptide sequence is defined as a polypeptide sequence having at least 40% sequence identity to the particular polypeptide sequence over a certain length of one of the polypeptide sequences using blastp with the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-1999) set at default parameters. Such a pair of polypeptides may show, for example, at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% or greater sequence identity over a certain defined length of one of the polypeptides.

THE INVENTION

The invention is based on the discovery of new human immunoglobulin superfamily proteins (IGSFP), the polynucleotides encoding IGSFP, and the use of these compositions for the diagnosis, treatment, or prevention of immune system, neurological, developmental, muscle, and cell proliferative disorders.

Table 1 summarizes the nomenclature for the full length polynucleotide and polypeptide sequences of the invention. Each polynucleotide and its corresponding polypeptide are correlated to a single Incyte project identification number (Incyte Project ID). Each polypeptide sequence is denoted by both a polypeptide sequence identification number (Polypeptide SEQ ID NO:) and an Incyte polypeptide sequence number (Incyte Polypeptide ID) as shown. Each polynucleotide sequence is denoted by both a polynucleotide sequence identification number (Polynucleotide SEQ ID NO:) and an Incyte polynucleotide consensus sequence number (Incyte Polynucleotide ID) as shown. Column 6 shows the Incyte ID numbers of physical, full length clones corresponding to the polypeptide and

polynucleotide sequences of the invention. The full length clones encode polypeptides which have at least 95% sequence identity to the polypeptide sequences shown in column 3.

Table 2 shows sequences with homology to the polypeptides of the invention as identified by BLAST analysis against the GenBank protein (genpept) database and the PROTEOME database. Columns 1 and 2 show the polypeptide sequence identification number (Polypeptide SEQ ID NO:) and the corresponding Incyte polypeptide sequence number (Incyte Polypeptide ID) for polypeptides of the invention. Column 3 shows the GenBank identification number (GenBank ID NO:) of the nearest GenBank homolog and the PROTEOME database identification numbers (PROTEOME ID NO:) of the nearest PROTEOME database homologs. Column 4 shows the probability scores for the matches between each polypeptide and its homolog(s). Column 5 shows the annotation of the GenBank and PROTEOME database homolog(s) along with relevant citations where applicable, all of which are expressly incorporated by reference herein.

Table 3 shows various structural features of the polypeptides of the invention. Columns 1 and 2 show the polypeptide sequence identification number (SEQ ID NO:) and the corresponding Incyte polypeptide sequence number (Incyte Polypeptide ID) for each polypeptide of the invention. Column 3 shows the number of amino acid residues in each polypeptide. Column 4 shows potential phosphorylation sites, and column 5 shows potential glycosylation sites, as determined by the MOTIFS program of the GCG sequence analysis software package (Genetics Computer Group, Madison WI). Column 6 shows amino acid residues comprising signature sequences, domains, and motifs. Column 7 shows analytical methods for protein structure/function analysis and in some cases, searchable databases to which the analytical methods were applied.

Together, Tables 2 and 3 summarize the properties of polypeptides of the invention, and these properties establish that the claimed polypeptides are immunoglobulin superfamily proteins. For example, SEQ ID NO:2 is 50% identical, from residue Q34 to residue P563, to Mus musculus Fca/m receptor (GenBank ID g11071950) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 9.6e-121, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:2 also contains an immunoglobulin domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) Data from additional BLAST analyses provide further corroborative evidence that SEQ ID NO:2 is an immunoglobulin. In an alternative example, SEQ ID NO:3 is 40% identical, from residue L30 to residue V176, to surface protein MCA-32 (GenBank ID g1136501) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 6.9e-35, which

indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:3 also contains an immunoglobulin domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) Data from BLIMPS, MOTIFS, and additional BLAST analyses provide further corroborative evidence that SEQ ID NO:3 is a surface protein. In an alternative example, SEQ ID NO:8 is 86% identical, from residue M1 to residue S433, to cell-surface molecule Ly-9 (GenBank ID g10197717) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 7.4e-191, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:8 also contains immunoglobulin domains as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) Data from additional BLAST analysis provide further corroborative evidence that SEQ ID NO:8 is a cell surface molecule which is a member of the immunoglobulin superfamily. In an alternative example, SEQ ID NO:11 is 52% identical, from residue N43 to residue Q604, to human NEPH1 (GenBank ID g14572521) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 5.4e-158, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. As determined by BLAST analysis using the PROTEOME database, SEQ ID NO:11 is localized to the plasma membrane, is homologous to a human protein which contains an immunoglobulin domain and has a region of low similarity to a region of an opioid-binding cell adhesion molecule, which is a glycosylphosphatidylinositol (GPI)-anchored neural cell adhesion molecule (PROTEOME ID 598720|FLJ10845); SEQ ID NO:11 is also homologous to human Nephrin which is a member of the immunoglobulin superfamily expressed in renal glomeruli which may have a role in the development or function of the kidney filtration barrier. Mutation of the Nephrin gene causes congenital nephrotic syndrome (PROTEOME ID 340970|NPHS1). SEQ ID NO:11 also contains an immunoglobulin domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) Data from BLIMPS, MOTIFS, and additional BLAST analyses provide further corroborative evidence that SEQ ID NO:11 is a member of the immunoglobulin superfamily. SEQ ID NO:1, SEQ ID NO:4-7, SEQ ID NO:9-10 and SEQ ID NO:12 were analyzed and annotated in a similar manner. The algorithms and parameters for the analysis of SEQ ID NO:1-12 are described in Table 7.

As shown in Table 4, the full length polynucleotide sequences of the present invention were assembled using cDNA sequences or coding (exon) sequences derived from genomic DNA, or any

combination of these two types of sequences. Column 1 lists the polynucleotide sequence identification number (Polynucleotide SEQ ID NO:), the corresponding Incyte polynucleotide consensus sequence number (Incyte ID) for each polynucleotide of the invention, and the length of each polynucleotide sequence in basepairs. Column 2 shows the nucleotide start (5') and stop (3') positions of the cDNA and/or genomic sequences used to assemble the full length polynucleotide sequences of the invention, and of fragments of the polynucleotide sequences which are useful, for example, in hybridization or amplification technologies that identify SEQ ID NO:13-24 or that distinguish between SEQ ID NO:13-24 and related polynucleotide sequences.

The polynucleotide fragments described in Column 2 of Table 4 may refer specifically, for example, to Incyte cDNAs derived from tissue-specific cDNA libraries or from pooled cDNA libraries. Alternatively, the polynucleotide fragments described in column 2 may refer to GenBank cDNAs or ESTs which contributed to the assembly of the full length polynucleotide sequences. In addition, the polynucleotide fragments described in column 2 may identify sequences derived from the ENSEMBL (The Sanger Centre, Cambridge, UK) database (*i.e.*, those sequences including the designation "ENST"). Alternatively, the polynucleotide fragments described in column 2 may be derived from the NCBI RefSeq Nucleotide Sequence Records Database (*i.e.*, those sequences including the designation "NM" or "NT") or the NCBI RefSeq Protein Sequence Records (*i.e.*, those sequences including the designation "NP"). Alternatively, the polynucleotide fragments described in column 2 may refer to assemblages of both cDNA and Genscan-predicted exons brought together by an "exon stitching" algorithm. For example, a polynucleotide sequence identified as FL_*XXXXXX*_N_1_N_2_*YYYYY*_N_3_N_4 represents a "stitched" sequence in which *XXXXXX* is the identification number of the cluster of sequences to which the algorithm was applied, and *YYYYY* is the number of the prediction generated by the algorithm, and $N_{1,2,3,..}$, if present, represent specific exons that may have been manually edited during analysis (See Example V). Alternatively, the polynucleotide fragments in column 2 may refer to assemblages of exons brought together by an "exon-stretching" algorithm. For example, a polynucleotide sequence identified as FL*XXXXXX*_g*AAAAA*_g*BBBBB*_1_*N* is a "stretched" sequence, with *XXXXXX* being the Incyte project identification number, g*AAAAA* being the GenBank identification number of the human genomic sequence to which the "exon-stretching" algorithm was applied, g*BBBBB* being the GenBank identification number or NCBI RefSeq identification number of the nearest GenBank protein homolog, and *N* referring to specific exons (See Example V). In instances where a RefSeq sequence was used as a protein homolog for the "exon-stretching" algorithm, a RefSeq identifier (denoted by "NM," "NP," or "NT") may be used in place of the GenBank identifier (*i.e.*, g*BBBBB*).

Alternatively, a prefix identifies component sequences that were hand-edited, predicted from genomic DNA sequences, or derived from a combination of sequence analysis methods. The following Table lists examples of component sequence prefixes and corresponding sequence analysis methods associated with the prefixes (see Example IV and Example V).

Prefix	Type of analysis and/or examples of programs
GNN, GFG, ENST	Exon prediction from genomic sequences using, for example, GENSCAN (Stanford University, CA, USA) or FGENES (Computer Genomics Group, The Sanger Centre, Cambridge, UK)
GBI	Hand-edited analysis of genomic sequences.
FL	Stitched or stretched genomic sequences (see Example V).
INCY	Full length transcript and exon prediction from mapping of EST sequences to the genome. Genomic location and EST composition data are combined to predict the exons and resulting transcript.

In some cases, Incyte cDNA coverage redundant with the sequence coverage shown in Table 4 was obtained to confirm the final consensus polynucleotide sequence, but the relevant Incyte cDNA identification numbers are not shown.

Table 5 shows the representative cDNA libraries for those full length polynucleotide sequences which were assembled using Incyte cDNA sequences. The representative cDNA library is the Incyte cDNA library which is most frequently represented by the Incyte cDNA sequences which were used to assemble and confirm the above polynucleotide sequences. The tissues and vectors which were used to construct the cDNA libraries shown in Table 5 are described in Table 6.

The invention also encompasses IGSFP variants. A preferred IGSFP variant is one which has at least about 80%, or alternatively at least about 90%, or even at least about 95% amino acid sequence identity to the IGSFP amino acid sequence, and which contains at least one functional or structural characteristic of IGSFP.

The invention also encompasses polynucleotides which encode IGSFP. In a particular embodiment, the invention encompasses a polynucleotide sequence comprising a sequence selected from the group consisting of SEQ ID NO:13-24, which encodes IGSFP. The polynucleotide sequences of SEQ ID NO:13-24, as presented in the Sequence Listing, embrace the equivalent RNA sequences, wherein occurrences of the nitrogenous base thymine are replaced with uracil, and the sugar backbone is composed of ribose instead of deoxyribose.

The invention also encompasses a variant of a polynucleotide sequence encoding IGSFP. In

particular, such a variant polynucleotide sequence will have at least about 70%, or alternatively at least about 85%, or even at least about 95% polynucleotide sequence identity to the polynucleotide sequence encoding IGSFP. A particular aspect of the invention encompasses a variant of a polynucleotide sequence comprising a sequence selected from the group consisting of SEQ ID NO:13-24 which has at least about 70%, or alternatively at least about 85%, or even at least about 95% polynucleotide sequence identity to a nucleic acid sequence selected from the group consisting of SEQ ID NO:13-24. Any one of the polynucleotide variants described above can encode an amino acid sequence which contains at least one functional or structural characteristic of IGSFP.

In addition, or in the alternative, a polynucleotide variant of the invention is a splice variant of a polynucleotide sequence encoding IGSFP. A splice variant may have portions which have significant sequence identity to the polynucleotide sequence encoding IGSFP, but will generally have a greater or lesser number of polynucleotides due to additions or deletions of blocks of sequence arising from alternate splicing of exons during mRNA processing. A splice variant may have less than about 70%, or alternatively less than about 60%, or alternatively less than about 50% polynucleotide sequence identity to the polynucleotide sequence encoding IGSFP over its entire length; however, portions of the splice variant will have at least about 70%, or alternatively at least about 85%, or alternatively at least about 95%, or alternatively 100% polynucleotide sequence identity to portions of the polynucleotide sequence encoding IGSFP. For example, a polynucleotide comprising a sequence of SEQ ID NO:14 is a splice variant of a polynucleotide comprising a sequence of SEQ ID NO:24 and a polynucleotide comprising a sequence of SEQ ID NO:16 is a splice variant of a polynucleotide comprising a sequence of SEQ ID NO:17. Any one of the splice variants described above can encode an amino acid sequence which contains at least one functional or structural characteristic of IGSFP.

It will be appreciated by those skilled in the art that as a result of the degeneracy of the genetic code, a multitude of polynucleotide sequences encoding IGSFP, some bearing minimal similarity to the polynucleotide sequences of any known and naturally occurring gene, may be produced. Thus, the invention contemplates each and every possible variation of polynucleotide sequence that could be made by selecting combinations based on possible codon choices. These combinations are made in accordance with the standard triplet genetic code as applied to the polynucleotide sequence of naturally occurring IGSFP, and all such variations are to be considered as being specifically disclosed.

Although nucleotide sequences which encode IGSFP and its variants are generally capable of hybridizing to the nucleotide sequence of the naturally occurring IGSFP under appropriately selected conditions of stringency, it may be advantageous to produce nucleotide sequences encoding IGSFP or

its derivatives possessing a substantially different codon usage, e.g., inclusion of non-naturally occurring codons. Codons may be selected to increase the rate at which expression of the peptide occurs in a particular prokaryotic or eukaryotic host in accordance with the frequency with which particular codons are utilized by the host. Other reasons for substantially altering the nucleotide sequence encoding IGSFP and its derivatives without altering the encoded amino acid sequences include the production of RNA transcripts having more desirable properties, such as a greater half-life, than transcripts produced from the naturally occurring sequence.

The invention also encompasses production of DNA sequences which encode IGSFP and IGSFP derivatives, or fragments thereof, entirely by synthetic chemistry. After production, the synthetic sequence may be inserted into any of the many available expression vectors and cell systems using reagents well known in the art. Moreover, synthetic chemistry may be used to introduce mutations into a sequence encoding IGSFP or any fragment thereof.

Also encompassed by the invention are polynucleotide sequences that are capable of hybridizing to the claimed polynucleotide sequences, and, in particular, to those shown in SEQ ID NO:13-24 and fragments thereof under various conditions of stringency. (See, e.g., Wahl, G.M. and S.L. Berger (1987) Methods Enzymol. 152:399-407; Kimmel, A.R. (1987) Methods Enzymol. 152:507-511.) Hybridization conditions, including annealing and wash conditions, are described in "Definitions."

Methods for DNA sequencing are well known in the art and may be used to practice any of the embodiments of the invention. The methods may employ such enzymes as the Klenow fragment of DNA polymerase I, SEQUENASE (US Biochemical, Cleveland OH), Taq polymerase (Applied Biosystems), thermostable T7 polymerase (Amersham Pharmacia Biotech, Piscataway NJ), or combinations of polymerases and proofreading exonucleases such as those found in the ELONGASE amplification system (Life Technologies, Gaithersburg MD). Preferably, sequence preparation is automated with machines such as the MICROLAB 2200 liquid transfer system (Hamilton, Reno NV), PTC200 thermal cycler (MJ Research, Watertown MA) and ABI CATALYST 800 thermal cycler (Applied Biosystems). Sequencing is then carried out using either the ABI 373 or 377 DNA sequencing system (Applied Biosystems), the MEGABACE 1000 DNA sequencing system (Molecular Dynamics, Sunnyvale CA), or other systems known in the art. The resulting sequences are analyzed using a variety of algorithms which are well known in the art. (See, e.g., Ausubel, F.M. (1997) Short Protocols in Molecular Biology, John Wiley & Sons, New York NY, unit 7.7; Meyers, R.A. (1995) Molecular Biology and Biotechnology, Wiley VCH, New York NY, pp. 856-853.)

The nucleic acid sequences encoding IGSFP may be extended utilizing a partial nucleotide sequence and employing various PCR-based methods known in the art to detect upstream sequences,

such as promoters and regulatory elements. For example, one method which may be employed, restriction-site PCR, uses universal and nested primers to amplify unknown sequence from genomic DNA within a cloning vector. (See, e.g., Sarkar, G. (1993) PCR Methods Applic. 2:318-322.) Another method, inverse PCR, uses primers that extend in divergent directions to amplify unknown sequence from a circularized template. The template is derived from restriction fragments comprising a known genomic locus and surrounding sequences. (See, e.g., Triglia, T. et al. (1988) Nucleic Acids Res. 16:8186.) A third method, capture PCR, involves PCR amplification of DNA fragments adjacent to known sequences in human and yeast artificial chromosome DNA. (See, e.g., Lagerstrom, M. et al. (1991) PCR Methods Applic. 1:111-119.) In this method, multiple restriction enzyme digestions and ligations may be used to insert an engineered double-stranded sequence into a region of unknown sequence before performing PCR. Other methods which may be used to retrieve unknown sequences are known in the art. (See, e.g., Parker, J.D. et al. (1991) Nucleic Acids Res. 19:3055-3060). Additionally, one may use PCR, nested primers, and PROMOTERFINDER libraries (Clontech, Palo Alto CA) to walk genomic DNA. This procedure avoids the need to screen libraries and is useful in finding intron/exon junctions. For all PCR-based methods, primers may be designed using commercially available software, such as OLIGO 4.06 primer analysis software (National Biosciences, Plymouth MN) or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the template at temperatures of about 68°C to 72°C.

When screening for full length cDNAs, it is preferable to use libraries that have been size-selected to include larger cDNAs. In addition, random-primed libraries, which often include sequences containing the 5' regions of genes, are preferable for situations in which an oligo d(T) library does not yield a full-length cDNA. Genomic libraries may be useful for extension of sequence into 5' non-transcribed regulatory regions.

Capillary electrophoresis systems which are commercially available may be used to analyze the size or confirm the nucleotide sequence of sequencing or PCR products. In particular, capillary sequencing may employ flowable polymers for electrophoretic separation, four different nucleotide-specific, laser-stimulated fluorescent dyes, and a charge coupled device camera for detection of the emitted wavelengths. Output/light intensity may be converted to electrical signal using appropriate software (e.g., GENOTYPER and SEQUENCE NAVIGATOR, Applied Biosystems), and the entire process from loading of samples to computer analysis and electronic data display may be computer controlled. Capillary electrophoresis is especially preferable for sequencing small DNA fragments which may be present in limited amounts in a particular sample.

In another embodiment of the invention, polynucleotide sequences or fragments thereof which encode IGSFP may be cloned in recombinant DNA molecules that direct expression of IGSFP, or fragments or functional equivalents thereof, in appropriate host cells. Due to the inherent degeneracy of the genetic code, other DNA sequences which encode substantially the same or a functionally equivalent amino acid sequence may be produced and used to express IGSFP.

The nucleotide sequences of the present invention can be engineered using methods generally known in the art in order to alter IGSFP-encoding sequences for a variety of purposes including, but not limited to, modification of the cloning, processing, and/or expression of the gene product. DNA shuffling by random fragmentation and PCR reassembly of gene fragments and synthetic oligonucleotides may be used to engineer the nucleotide sequences. For example, oligonucleotide-mediated site-directed mutagenesis may be used to introduce mutations that create new restriction sites, alter glycosylation patterns, change codon preference, produce splice variants, and so forth.

The nucleotides of the present invention may be subjected to DNA shuffling techniques such as MOLECULARBREEDING (Maxygen Inc., Santa Clara CA; described in U.S. Patent No. 5,837,458; Chang, C.-C. et al. (1999) Nat. Biotechnol. 17:793-797; Christians, F.C. et al. (1999) Nat. Biotechnol. 17:259-264; and Crameri, A. et al. (1996) Nat. Biotechnol. 14:315-319) to alter or improve the biological properties of IGSFP, such as its biological or enzymatic activity or its ability to bind to other molecules or compounds. DNA shuffling is a process by which a library of gene variants is produced using PCR-mediated recombination of gene fragments. The library is then subjected to selection or screening procedures that identify those gene variants with the desired properties. These preferred variants may then be pooled and further subjected to recursive rounds of DNA shuffling and selection/screening. Thus, genetic diversity is created through "artificial" breeding and rapid molecular evolution. For example, fragments of a single gene containing random point mutations may be recombined, screened, and then reshuffled until the desired properties are optimized. Alternatively, fragments of a given gene may be recombined with fragments of homologous genes in the same gene family, either from the same or different species, thereby maximizing the genetic diversity of multiple naturally occurring genes in a directed and controllable manner.

In another embodiment, sequences encoding IGSFP may be synthesized, in whole or in part, using chemical methods well known in the art. (See, e.g., Caruthers, M.H. et al. (1980) Nucleic Acids Symp. Ser. 7:215-223; and Horn, T. et al. (1980) Nucleic Acids Symp. Ser. 7:225-232.) Alternatively, IGSFP itself or a fragment thereof may be synthesized using chemical methods. For example, peptide synthesis can be performed using various solution-phase or solid-phase techniques. (See, e.g., Creighton, T. (1984) Proteins, Structures and Molecular Properties, WH Freeman, New York NY, pp.

55-60; and Roberge, J.Y. et al. (1995) Science 269:202-204.) Automated synthesis may be achieved using the ABI 431A peptide synthesizer (Applied Biosystems). Additionally, the amino acid sequence of IGSFP, or any part thereof, may be altered during direct synthesis and/or combined with sequences from other proteins, or any part thereof, to produce a variant polypeptide or a polypeptide having a sequence of a naturally occurring polypeptide.

The peptide may be substantially purified by preparative high performance liquid chromatography. (See, e.g., Chiez, R.M. and F.Z. Regnier (1990) Methods Enzymol. 182:392-421.) The composition of the synthetic peptides may be confirmed by amino acid analysis or by sequencing. (See, e.g., Creighton, supra, pp. 28-53.)

In order to express a biologically active IGSFP, the nucleotide sequences encoding IGSFP or derivatives thereof may be inserted into an appropriate expression vector, i.e., a vector which contains the necessary elements for transcriptional and translational control of the inserted coding sequence in a suitable host. These elements include regulatory sequences, such as enhancers, constitutive and inducible promoters, and 5' and 3' untranslated regions in the vector and in polynucleotide sequences encoding IGSFP. Such elements may vary in their strength and specificity. Specific initiation signals may also be used to achieve more efficient translation of sequences encoding IGSFP. Such signals include the ATG initiation codon and adjacent sequences, e.g. the Kozak sequence. In cases where sequences encoding IGSFP and its initiation codon and upstream regulatory sequences are inserted into the appropriate expression vector, no additional transcriptional or translational control signals may be needed. However, in cases where only coding sequence, or a fragment thereof, is inserted, exogenous translational control signals including an in-frame ATG initiation codon should be provided by the vector. Exogenous translational elements and initiation codons may be of various origins, both natural and synthetic. The efficiency of expression may be enhanced by the inclusion of enhancers appropriate for the particular host cell system used. (See, e.g., Scharf, D. et al. (1994) Results Probl. Cell Differ. 20:125-162.)

Methods which are well known to those skilled in the art may be used to construct expression vectors containing sequences encoding IGSFP and appropriate transcriptional and translational control elements. These methods include in vitro recombinant DNA techniques, synthetic techniques, and in vivo genetic recombination. (See, e.g., Sambrook, J. et al. (1989) Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Press, Plainview NY, ch. 4, 8, and 16-17; Ausubel, F.M. et al. (1995) Current Protocols in Molecular Biology, John Wiley & Sons, New York NY, ch. 9, 13, and 16.)

A variety of expression vector/host systems may be utilized to contain and express sequences encoding IGSFP. These include, but are not limited to, microorganisms such as bacteria transformed

with recombinant bacteriophage, plasmid, or cosmid DNA expression vectors; yeast transformed with yeast expression vectors; insect cell systems infected with viral expression vectors (e.g., baculovirus); plant cell systems transformed with viral expression vectors (e.g., cauliflower mosaic virus, CaMV, or tobacco mosaic virus, TMV) or with bacterial expression vectors (e.g., Ti or pBR322 plasmids); or animal cell systems. (See, e.g., Sambrook, supra; Ausubel, supra; Van Heeke, G. and S.M. Schuster (1989) J. Biol. Chem. 264:5503-5509; Engelhard, E.K. et al. (1994) Proc. Natl. Acad. Sci. USA 91:3224-3227; Sandig, V. et al. (1996) Hum. Gene Ther. 7:1937-1945; Takamatsu, N. (1987) EMBO J. 6:307-311; The McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill, New York NY, pp. 191-196; Logan, J. and T. Shenk (1984) Proc. Natl. Acad. Sci. USA 81:3655-3659; and Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355.) Expression vectors derived from retroviruses, adenoviruses, or herpes or vaccinia viruses, or from various bacterial plasmids, may be used for delivery of nucleotide sequences to the targeted organ, tissue, or cell population. (See, e.g., Di Nicola, M. et al. (1998) Cancer Gen. Ther. 5(6):350-356; Yu, M. et al. (1993) Proc. Natl. Acad. Sci. USA 90(13):6340-6344; Buller, R.M. et al. (1985) Nature 317(6040):813-815; McGregor, D.P. et al. (1994) Mol. Immunol. 31(3):219-226; and Verma, I.M. and N. Somia (1997) Nature 389:239-242.) The invention is not limited by the host cell employed.

In bacterial systems, a number of cloning and expression vectors may be selected depending upon the use intended for polynucleotide sequences encoding IGSFP. For example, routine cloning, subcloning, and propagation of polynucleotide sequences encoding IGSFP can be achieved using a multifunctional E. coli vector such as PBLUESCRIPT (Stratagene, La Jolla CA) or PSPORT1 plasmid (Life Technologies). Ligation of sequences encoding IGSFP into the vector's multiple cloning site disrupts the *lac*Z gene, allowing a colorimetric screening procedure for identification of transformed bacteria containing recombinant molecules. In addition, these vectors may be useful for in vitro transcription, dideoxy sequencing, single strand rescue with helper phage, and creation of nested deletions in the cloned sequence. (See, e.g., Van Heeke, G. and S.M. Schuster (1989) J. Biol. Chem. 264:5503-5509.) When large quantities of IGSFP are needed, e.g. for the production of antibodies, vectors which direct high level expression of IGSFP may be used. For example, vectors containing the strong, inducible SP6 or T7 bacteriophage promoter may be used.

Yeast expression systems may be used for production of IGSFP. A number of vectors containing constitutive or inducible promoters, such as alpha factor, alcohol oxidase, and PGH promoters, may be used in the yeast Saccharomyces cerevisiae or Pichia pastoris. In addition, such vectors direct either the secretion or intracellular retention of expressed proteins and enable integration of foreign sequences into the host genome for stable propagation. (See, e.g., Ausubel, 1995, supra;

Bitter, G.A. et al. (1987) Methods Enzymol. 153:516-544; and Scorer, C.A. et al. (1994) Bio/Technology 12:181-184.)

Plant systems may also be used for expression of IGSFP. Transcription of sequences encoding IGSFP may be driven by viral promoters, e.g., the 35S and 19S promoters of CaMV used alone or in combination with the omega leader sequence from TMV (Takamatsu, N. (1987) EMBO J. 6:307-311). Alternatively, plant promoters such as the small subunit of RUBISCO or heat shock promoters may be used. (See, e.g., Coruzzi, G. et al. (1984) EMBO J. 3:1671-1680; Broglie, R. et al. (1984) Science 224:838-843; and Winter, J. et al. (1991) Results Probl. Cell Differ. 17:85-105.) These constructs can be introduced into plant cells by direct DNA transformation or pathogen-mediated transfection. (See, e.g., The McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill, New York NY, pp. 191-196.)

In mammalian cells, a number of viral-based expression systems may be utilized. In cases where an adenovirus is used as an expression vector, sequences encoding IGSFP may be ligated into an adenovirus transcription/translation complex consisting of the late promoter and tripartite leader sequence. Insertion in a non-essential E1 or E3 region of the viral genome may be used to obtain infective virus which expresses IGSFP in host cells. (See, e.g., Logan, J. and T. Shenk (1984) Proc. Natl. Acad. Sci. USA 81:3655-3659.) In addition, transcription enhancers, such as the Rous sarcoma virus (RSV) enhancer, may be used to increase expression in mammalian host cells. SV40 or EBV-based vectors may also be used for high-level protein expression.

Human artificial chromosomes (HACs) may also be employed to deliver larger fragments of DNA than can be contained in and expressed from a plasmid. HACs of about 6 kb to 10 Mb are constructed and delivered via conventional delivery methods (liposomes, polycationic amino polymers, or vesicles) for therapeutic purposes. (See, e.g., Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355.)

For long term production of recombinant proteins in mammalian systems, stable expression of IGSFP in cell lines is preferred. For example, sequences encoding IGSFP can be transformed into cell lines using expression vectors which may contain viral origins of replication and/or endogenous expression elements and a selectable marker gene on the same or on a separate vector. Following the introduction of the vector, cells may be allowed to grow for about 1 to 2 days in enriched media before being switched to selective media. The purpose of the selectable marker is to confer resistance to a selective agent, and its presence allows growth and recovery of cells which successfully express the introduced sequences. Resistant clones of stably transformed cells may be propagated using tissue culture techniques appropriate to the cell type.

Any number of selection systems may be used to recover transformed cell lines. These include, but are not limited to, the herpes simplex virus thymidine kinase and adenine phosphoribosyltransferase genes, for use in *tk* and *apr* cells, respectively. (See, e.g., Wigler, M. et al. (1977) Cell 11:223-232; Lowy, I. et al. (1980) Cell 22:817-823.) Also, antimetabolite, antibiotic, or herbicide resistance can be used as the basis for selection. For example, *dhfr* confers resistance to methotrexate; *neo* confers resistance to the aminoglycosides neomycin and G-418; and *als* and *pat* confer resistance to chlorsulfuron and phosphinotricin acetyltransferase, respectively. (See, e.g., Wigler, M. et al. (1980) Proc. Natl. Acad. Sci. USA 77:3567-3570; Colbere-Garapin, F. et al. (1981) J. Mol. Biol. 150:1-14.) Additional selectable genes have been described, e.g., *trpB* and *hisD*, which alter cellular requirements for metabolites. (See, e.g., Hartman, S.C. and R.C. Mulligan (1988) Proc. Natl. Acad. Sci. USA 85:8047-8051.) Visible markers, e.g., anthocyanins, green fluorescent proteins (GFP; Clontech), ß glucuronidase and its substrate ß-glucuronide, or luciferase and its substrate luciferin may be used. These markers can be used not only to identify transformants, but also to quantify the amount of transient or stable protein expression attributable to a specific vector system. (See, e.g., Rhodes, C.A. (1995) Methods Mol. Biol. 55:121-131.)

Although the presence/absence of marker gene expression suggests that the gene of interest is also present, the presence and expression of the gene may need to be confirmed. For example, if the sequence encoding IGSFP is inserted within a marker gene sequence, transformed cells containing sequences encoding IGSFP can be identified by the absence of marker gene function. Alternatively, a marker gene can be placed in tandem with a sequence encoding IGSFP under the control of a single promoter. Expression of the marker gene in response to induction or selection usually indicates expression of the tandem gene as well.

In general, host cells that contain the nucleic acid sequence encoding IGSFP and that express IGSFP may be identified by a variety of procedures known to those of skill in the art. These procedures include, but are not limited to, DNA-DNA or DNA-RNA hybridizations, PCR amplification, and protein bioassay or immunoassay techniques which include membrane, solution, or chip based technologies for the detection and/or quantification of nucleic acid or protein sequences.

Immunological methods for detecting and measuring the expression of IGSFP using either specific polyclonal or monoclonal antibodies are known in the art. Examples of such techniques include enzyme-linked immunosorbent assays (ELISAs), radioimmunoassays (RIAs), and fluorescence activated cell sorting (FACS). A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering epitopes on IGSFP is preferred, but a competitive binding assay may be employed. These and other assays are well known in the art. (See,

e.g., Hampton, R. et al. (1990) Serological Methods, a Laboratory Manual, APS Press, St. Paul MN, Sect. IV; Coligan, J.E. et al. (1997) Current Protocols in Immunology, Greene Pub. Associates and Wiley-Interscience, New York NY; and Pound, J.D. (1998) Immunochemical Protocols, Humana Press, Totowa NJ.)

A wide variety of labels and conjugation techniques are known by those skilled in the art and may be used in various nucleic acid and amino acid assays. Means for producing labeled hybridization or PCR probes for detecting sequences related to polynucleotides encoding IGSFP include oligolabeling, nick translation, end-labeling, or PCR amplification using a labeled nucleotide. Alternatively, the sequences encoding IGSFP, or any fragments thereof, may be cloned into a vector for the production of an mRNA probe. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes in vitro by addition of an appropriate RNA polymerase such as T7, T3, or SP6 and labeled nucleotides. These procedures may be conducted using a variety of commercially available kits, such as those provided by Amersham Pharmacia Biotech, Promega (Madison WI), and US Biochemical. Suitable reporter molecules or labels which may be used for ease of detection include radionuclides, enzymes, fluorescent, chemiluminescent, or chromogenic agents, as well as substrates, cofactors, inhibitors, magnetic particles, and the like.

Host cells transformed with nucleotide sequences encoding IGSFP may be cultured under conditions suitable for the expression and recovery of the protein from cell culture. The protein produced by a transformed cell may be secreted or retained intracellularly depending on the sequence and/or the vector used. As will be understood by those of skill in the art, expression vectors containing polynucleotides which encode IGSFP may be designed to contain signal sequences which direct secretion of IGSFP through a prokaryotic or eukaryotic cell membrane.

In addition, a host cell strain may be chosen for its ability to modulate expression of the inserted sequences or to process the expressed protein in the desired fashion. Such modifications of the polypeptide include, but are not limited to, acetylation, carboxylation, glycosylation, phosphorylation, lipidation, and acylation. Post-translational processing which cleaves a "prepro" or "pro" form of the protein may also be used to specify protein targeting, folding, and/or activity. Different host cells which have specific cellular machinery and characteristic mechanisms for post-translational activities (e.g., CHO, HeLa, MDCK, HEK293, and WI38) are available from the American Type Culture Collection (ATCC, Manassas VA) and may be chosen to ensure the correct modification and processing of the foreign protein.

In another embodiment of the invention, natural, modified, or recombinant nucleic acid sequences encoding IGSFP may be ligated to a heterologous sequence resulting in translation of a

fusion protein in any of the aforementioned host systems. For example, a chimeric IGSFP protein containing a heterologous moiety that can be recognized by a commercially available antibody may facilitate the screening of peptide libraries for inhibitors of IGSFP activity. Heterologous protein and peptide moieties may also facilitate purification of fusion proteins using commercially available affinity matrices. Such moieties include, but are not limited to, glutathione S-transferase (GST), maltose binding protein (MBP), thioredoxin (Trx), calmodulin binding peptide (CBP), 6-His, FLAG, *c-myc*, and hemagglutinin (HA). GST, MBP, Trx, CBP, and 6-His enable purification of their cognate fusion proteins on immobilized glutathione, maltose, phenylarsine oxide, calmodulin, and metal-chelate resins, respectively. FLAG, *c-myc*, and hemagglutinin (HA) enable immunoaffinity purification of fusion proteins using commercially available monoclonal and polyclonal antibodies that specifically recognize these epitope tags. A fusion protein may also be engineered to contain a proteolytic cleavage site located between the IGSFP encoding sequence and the heterologous protein sequence, so that IGSFP may be cleaved away from the heterologous moiety following purification. Methods for fusion protein expression and purification are discussed in Ausubel (1995, supra, ch. 10). A variety of commercially available kits may also be used to facilitate expression and purification of fusion proteins.

In a further embodiment of the invention, synthesis of radiolabeled IGSFP may be achieved in vitro using the TNT rabbit reticulocyte lysate or wheat germ extract system (Promega). These systems couple transcription and translation of protein-coding sequences operably associated with the T7, T3, or SP6 promoters. Translation takes place in the presence of a radiolabeled amino acid precursor, for example, ^{35}S-methionine.

IGSFP of the present invention or fragments thereof may be used to screen for compounds that specifically bind to IGSFP. At least one and up to a plurality of test compounds may be screened for specific binding to IGSFP. Examples of test compounds include antibodies, oligonucleotides, proteins (e.g., receptors), or small molecules.

In one embodiment, the compound thus identified is closely related to the natural ligand of IGSFP, e.g., a ligand or fragment thereof, a natural substrate, a structural or functional mimetic, or a natural binding partner. (See, e.g., Coligan, J.E. et al. (1991) Current Protocols in Immunology 1(2): Chapter 5.) Similarly, the compound can be closely related to the natural receptor to which IGSFP binds, or to at least a fragment of the receptor, e.g., the ligand binding site. In either case, the compound can be rationally designed using known techniques. In one embodiment, screening for these compounds involves producing appropriate cells which express IGSFP, either as a secreted protein or on the cell membrane. Preferred cells include cells from mammals, yeast, Drosophila, or E. coli. Cells expressing IGSFP or cell membrane fractions which contain IGSFP are then contacted

with a test compound and binding, stimulation, or inhibition of activity of either IGSFP or the compound is analyzed.

An assay may simply test binding of a test compound to the polypeptide, wherein binding is detected by a fluorophore, radioisotope, enzyme conjugate, or other detectable label. For example, the assay may comprise the steps of combining at least one test compound with IGSFP, either in solution or affixed to a solid support, and detecting the binding of IGSFP to the compound. Alternatively, the assay may detect or measure binding of a test compound in the presence of a labeled competitor. Additionally, the assay may be carried out using cell-free preparations, chemical libraries, or natural product mixtures, and the test compound(s) may be free in solution or affixed to a solid support.

IGSFP of the present invention or fragments thereof may be used to screen for compounds that modulate the activity of IGSFP. Such compounds may include agonists, antagonists, or partial or inverse agonists. In one embodiment, an assay is performed under conditions permissive for IGSFP activity, wherein IGSFP is combined with at least one test compound, and the activity of IGSFP in the presence of a test compound is compared with the activity of IGSFP in the absence of the test compound. A change in the activity of IGSFP in the presence of the test compound is indicative of a compound that modulates the activity of IGSFP. Alternatively, a test compound is combined with an in vitro or cell-free system comprising IGSFP under conditions suitable for IGSFP activity, and the assay is performed. In either of these assays, a test compound which modulates the activity of IGSFP may do so indirectly and need not come in direct contact with the test compound. At least one and up to a plurality of test compounds may be screened.

In another embodiment, polynucleotides encoding IGSFP or their mammalian homologs may be "knocked out" in an animal model system using homologous recombination in embryonic stem (ES) cells. Such techniques are well known in the art and are useful for the generation of animal models of human disease. (See, e.g., U.S. Patent No. 5,175,383 and U.S. Patent No. 5,767,337.) For example, mouse ES cells, such as the mouse 129/SvJ cell line, are derived from the early mouse embryo and grown in culture. The ES cells are transformed with a vector containing the gene of interest disrupted by a marker gene, e.g., the neomycin phosphotransferase gene (neo; Capecchi, M.R. (1989) Science 244:1288-1292). The vector integrates into the corresponding region of the host genome by homologous recombination. Alternatively, homologous recombination takes place using the Cre-loxP system to knockout a gene of interest in a tissue- or developmental stage-specific manner (Marth, J.D. (1996) Clin. Invest. 97:1999-2002; Wagner, K.U. et al. (1997) Nucleic Acids Res. 25:4323-4330). Transformed ES cells are identified and microinjected into mouse cell blastocysts such as those from the C57BL/6 mouse strain. The blastocysts are surgically transferred to pseudopregnant dams, and

the resulting chimeric progeny are genotyped and bred to produce heterozygous or homozygous strains. Transgenic animals thus generated may be tested with potential therapeutic or toxic agents.

Polynucleotides encoding IGSFP may also be manipulated in vitro in ES cells derived from human blastocysts. Human ES cells have the potential to differentiate into at least eight separate cell lineages including endoderm, mesoderm, and ectodermal cell types. These cell lineages differentiate into, for example, neural cells, hematopoietic lineages, and cardiomyocytes (Thomson, J.A. et al. (1998) Science 282:1145-1147).

Polynucleotides encoding IGSFP can also be used to create "knockin" humanized animals (pigs) or transgenic animals (mice or rats) to model human disease. With knockin technology, a region of a polynucleotide encoding IGSFP is injected into animal ES cells, and the injected sequence integrates into the animal cell genome. Transformed cells are injected into blastulae, and the blastulae are implanted as described above. Transgenic progeny or inbred lines are studied and treated with potential pharmaceutical agents to obtain information on treatment of a human disease. Alternatively, a mammal inbred to overexpress IGSFP, e.g., by secreting IGSFP in its milk, may also serve as a convenient source of that protein (Janne, J. et al. (1998) Biotechnol. Annu. Rev. 4:55-74).

THERAPEUTICS

Chemical and structural similarity, e.g., in the context of sequences and motifs, exists between regions of IGSFP and immunoglobulin superfamily proteins. In addition, the expression of IGSFP is closely associated with brain, colon, diseased skin, diseased lung, hippocampus, spleen, and diseased vermis tissues, as well as, $CD4^+$ T and peripheral blood cells. Therefore, IGSFP appears to play a role in immune system, neurological, developmental, muscle, and cell proliferative disorders. In the treatment of disorders associated with increased IGSFP expression or activity, it is desirable to decrease the expression or activity of IGSFP. In the treatment of disorders associated with decreased IGSFP expression or activity, it is desirable to increase the expression or activity of IGSFP.

Therefore, in one embodiment, IGSFP or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of IGSFP. Examples of such disorders include, but are not limited to, an immune system disorder such as acquired immunodeficiency syndrome (AIDS), X-linked agammaglobinemia of Bruton, common variable immunodeficiency (CVI), DiGeorge's syndrome (thymic hypoplasia), thymic dysplasia, isolated IgA deficiency, severe combined immunodeficiency disease (SCID), immunodeficiency with thrombocytopenia and eczema (Wiskott-Aldrich syndrome), Chediak-Higashi syndrome, chronic granulomatous diseases, hereditary angioneurotic edema, immunodeficiency associated with Cushing's disease, Addison's disease, adult respiratory distress syndrome, allergies,

ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis, autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED), bronchitis, cholecystitis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, episodic lymphopenia with lymphocytotoxins, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, hypereosinophilia, irritable bowel syndrome, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, viral, bacterial, fungal, parasitic, protozoal, and helminthic infections, and trauma; a neurological disorder such as epilepsy, ischemic cerebrovascular disease, stroke, cerebral neoplasms, Alzheimer's disease, Pick's disease, Huntington's disease, dementia, Parkinson's disease and other extrapyramidal disorders, amyotrophic lateral sclerosis and other motor neuron disorders, progressive neural muscular atrophy, retinitis pigmentosa, hereditary ataxias, multiple sclerosis and other demyelinating diseases, bacterial and viral meningitis, brain abscess, subdural empyema, epidural abscess, suppurative intracranial thrombophlebitis, myelitis and radiculitis, viral central nervous system disease, prion diseases including kuru, Creutzfeldt-Jakob disease, and Gerstmann-Straussler-Scheinker syndrome, fatal familial insomnia, nutritional and metabolic diseases of the nervous system, neurofibromatosis, tuberous sclerosis, cerebelloretinal hemangioblastomatosis, encephalotrigeminal syndrome, mental retardation and other developmental disorders of the central nervous system including Down syndrome, cerebral palsy, neuroskeletal disorders, autonomic nervous system disorders, cranial nerve disorders, spinal cord diseases, muscular dystrophy and other neuromuscular disorders, peripheral nervous system disorders, dermatomyositis and polymyositis, inherited, metabolic, endocrine, and toxic myopathies, myasthenia gravis, periodic paralysis, mental disorders including mood, anxiety, and schizophrenic disorders, seasonal affective disorder (SAD), akathesia, amnesia, catatonia, diabetic neuropathy, tardive dyskinesia, dystonias, paranoid psychoses, postherpetic neuralgia, Tourette's disorder, progressive supranuclear palsy, corticobasal degeneration, and familial frontotemporal dementia; a developmental disorder such as renal tubular acidosis, anemia, Cushing's syndrome, achondroplastic dwarfism, Duchenne and Becker muscular dystrophy, epilepsy, gonadal dysgenesis, WAGR syndrome (Wilms' tumor, aniridia, genitourinary abnormalities, and mental retardation), Smith-Magenis syndrome, myelodysplastic syndrome, hereditary mucoepithelial dysplasia, hereditary keratodermas, hereditary neuropathies such as Charcot-Marie-Tooth disease and

neurofibromatosis, hypothyroidism, hydrocephalus, seizure disorders such as Syndenham's chorea and cerebral palsy, spina bifida, anencephaly, craniorachischisis, congenital glaucoma, cataract, and sensorineural hearing loss; a muscle disorder such as cardiomyopathy, myocarditis, Duchenne's muscular dystrophy, Becker's muscular dystrophy, myotonic dystrophy, central core disease, nemaline myopathy, centronuclear myopathy, lipid myopathy, mitochondrial myopathy, infectious myositis, polymyositis, dermatomyositis, inclusion body myositis, thyrotoxic myopathy, and ethanol myopathy; and a cell proliferative disorder such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, and cancers including adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus.

In another embodiment, a vector capable of expressing IGSFP or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of IGSFP including, but not limited to, those described above.

In a further embodiment, a composition comprising a substantially purified IGSFP in conjunction with a suitable pharmaceutical carrier may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of IGSFP including, but not limited to, those provided above.

In still another embodiment, an agonist which modulates the activity of IGSFP may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of IGSFP including, but not limited to, those listed above.

In a further embodiment, an antagonist of IGSFP may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of IGSFP. Examples of such disorders include, but are not limited to, those immune system, neurological, developmental, muscle, and cell proliferative disorders described above. In one aspect, an antibody which specifically binds IGSFP may be used directly as an antagonist or indirectly as a targeting or delivery mechanism for bringing a pharmaceutical agent to cells or tissues which express IGSFP.

In an additional embodiment, a vector expressing the complement of the polynucleotide encoding IGSFP may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of IGSFP including, but not limited to, those described above.

In other embodiments, any of the proteins, antagonists, antibodies, agonists, complementary

sequences, or vectors of the invention may be administered in combination with other appropriate therapeutic agents. Selection of the appropriate agents for use in combination therapy may be made by one of ordinary skill in the art, according to conventional pharmaceutical principles. The combination of therapeutic agents may act synergistically to effect the treatment or prevention of the various disorders described above. Using this approach, one may be able to achieve therapeutic efficacy with lower dosages of each agent, thus reducing the potential for adverse side effects.

An antagonist of IGSFP may be produced using methods which are generally known in the art. In particular, purified IGSFP may be used to produce antibodies or to screen libraries of pharmaceutical agents to identify those which specifically bind IGSFP. Antibodies to IGSFP may also be generated using methods that are well known in the art. Such antibodies may include, but are not limited to, polyclonal, monoclonal, chimeric, and single chain antibodies, Fab fragments, and fragments produced by a Fab expression library. Neutralizing antibodies (i.e., those which inhibit dimer formation) are generally preferred for therapeutic use. Single chain antibodies (e.g., from camels or llamas) may be potent enzyme inhibitors and may have advantages in the design of peptide mimetics, and in the development of immuno-adsorbents and biosensors (Muyldermans, S. (2001) J. Biotechnol. 74:277-302).

For the production of antibodies, various hosts including goats, rabbits, rats, mice, camels, dromedaries, llamas, humans, and others may be immunized by injection with IGSFP or with any fragment or oligopeptide thereof which has immunogenic properties. Depending on the host species, various adjuvants may be used to increase immunological response. Such adjuvants include, but are not limited to, Freund's, mineral gels such as aluminum hydroxide, and surface active substances such as lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, KLH, and dinitrophenol. Among adjuvants used in humans, BCG (bacilli Calmette-Guerin) and Corynebacterium parvum are especially preferable.

It is preferred that the oligopeptides, peptides, or fragments used to induce antibodies to IGSFP have an amino acid sequence consisting of at least about 5 amino acids, and generally will consist of at least about 10 amino acids. It is also preferable that these oligopeptides, peptides, or fragments are identical to a portion of the amino acid sequence of the natural protein. Short stretches of IGSFP amino acids may be fused with those of another protein, such as KLH, and antibodies to the chimeric molecule may be produced.

Monoclonal antibodies to IGSFP may be prepared using any technique which provides for the production of antibody molecules by continuous cell lines in culture. These include, but are not limited to, the hybridoma technique, the human B-cell hybridoma technique, and the EBV-hybridoma

technique. (See, e.g., Kohler, G. et al. (1975) Nature 256:495-497; Kozbor, D. et al. (1985) J. Immunol. Methods 81:31-42; Cote, R.J. et al. (1983) Proc. Natl. Acad. Sci. USA 80:2026-2030; and Cole, S.P. et al. (1984) Mol. Cell Biol. 62:109-120.)

In addition, techniques developed for the production of "chimeric antibodies," such as the splicing of mouse antibody genes to human antibody genes to obtain a molecule with appropriate antigen specificity and biological activity, can be used. (See, e.g., Morrison, S.L. et al. (1984) Proc. Natl. Acad. Sci. USA 81:6851-6855; Neuberger, M.S. et al. (1984) Nature 312:604-608; and Takeda, S. et al. (1985) Nature 314:452-454.) Alternatively, techniques described for the production of single chain antibodies may be adapted, using methods known in the art, to produce IGSFP-specific single chain antibodies. Antibodies with related specificity, but of distinct idiotypic composition, may be generated by chain shuffling from random combinatorial immunoglobulin libraries. (See, e.g., Burton, D.R. (1991) Proc. Natl. Acad. Sci. USA 88:10134-10137.)

Antibodies may also be produced by inducing in vivo production in the lymphocyte population or by screening immunoglobulin libraries or panels of highly specific binding reagents as disclosed in the literature. (See, e.g., Orlandi, R. et al. (1989) Proc. Natl. Acad. Sci. USA 86:3833-3837; Winter, G. et al. (1991) Nature 349:293-299.)

Antibody fragments which contain specific binding sites for IGSFP may also be generated. For example, such fragments include, but are not limited to, $F(ab')_2$ fragments produced by pepsin digestion of the antibody molecule and Fab fragments generated by reducing the disulfide bridges of the F(ab')2 fragments. Alternatively, Fab expression libraries may be constructed to allow rapid and easy identification of monoclonal Fab fragments with the desired specificity. (See, e.g., Huse, W.D. et al. (1989) Science 246:1275-1281.)

Various immunoassays may be used for screening to identify antibodies having the desired specificity. Numerous protocols for competitive binding or immunoradiometric assays using either polyclonal or monoclonal antibodies with established specificities are well known in the art. Such immunoassays typically involve the measurement of complex formation between IGSFP and its specific antibody. A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering IGSFP epitopes is generally used, but a competitive binding assay may also be employed (Pound, supra).

Various methods such as Scatchard analysis in conjunction with radioimmunoassay techniques may be used to assess the affinity of antibodies for IGSFP. Affinity is expressed as an association constant, K_a, which is defined as the molar concentration of IGSFP-antibody complex divided by the molar concentrations of free antigen and free antibody under equilibrium conditions. The K_a

determined for a preparation of polyclonal antibodies, which are heterogeneous in their affinities for multiple IGSFP epitopes, represents the average affinity, or avidity, of the antibodies for IGSFP. The K_a determined for a preparation of monoclonal antibodies, which are monospecific for a particular IGSFP epitope, represents a true measure of affinity. High-affinity antibody preparations with K_a ranging from about 10^9 to 10^{12} L/mole are preferred for use in immunoassays in which the IGSFP-antibody complex must withstand rigorous manipulations. Low-affinity antibody preparations with K_a ranging from about 10^6 to 10^7 L/mole are preferred for use in immunopurification and similar procedures which ultimately require dissociation of IGSFP, preferably in active form, from the antibody (Catty, D. (1988) Antibodies, Volume I: A Practical Approach, IRL Press, Washington DC; Liddell, J.E. and A. Cryer (1991) A Practical Guide to Monoclonal Antibodies, John Wiley & Sons, New York NY).

The titer and avidity of polyclonal antibody preparations may be further evaluated to determine the quality and suitability of such preparations for certain downstream applications. For example, a polyclonal antibody preparation containing at least 1-2 mg specific antibody/ml, preferably 5-10 mg specific antibody/ml, is generally employed in procedures requiring precipitation of IGSFP-antibody complexes. Procedures for evaluating antibody specificity, titer, and avidity, and guidelines for antibody quality and usage in various applications, are generally available. (See, e.g., Catty, supra, and Coligan et al. supra.)

In another embodiment of the invention, the polynucleotides encoding IGSFP, or any fragment or complement thereof, may be used for therapeutic purposes. In one aspect, modifications of gene expression can be achieved by designing complementary sequences or antisense molecules (DNA, RNA, PNA, or modified oligonucleotides) to the coding or regulatory regions of the gene encoding IGSFP. Such technology is well known in the art, and antisense oligonucleotides or larger fragments can be designed from various locations along the coding or control regions of sequences encoding IGSFP. (See, e.g., Agrawal, S., ed. (1996) Antisense Therapeutics, Humana Press Inc., Totawa NJ.)

In therapeutic use, any gene delivery system suitable for introduction of the antisense sequences into appropriate target cells can be used. Antisense sequences can be delivered intracellularly in the form of an expression plasmid which, upon transcription, produces a sequence complementary to at least a portion of the cellular sequence encoding the target protein. (See, e.g., Slater, J.E. et al. (1998) J. Allergy Clin. Immunol. 102(3):469-475; and Scanlon, K.J. et al. (1995) 9(13):1288-1296.) Antisense sequences can also be introduced intracellularly through the use of viral vectors, such as retrovirus and adeno-associated virus vectors. (See, e.g., Miller, A.D. (1990) Blood 76:271; Ausubel, supra; Uckert, W. and W. Walther (1994) Pharmacol. Ther. 63(3):323-347.) Other

gene delivery mechanisms include liposome-derived systems, artificial viral envelopes, and other systems known in the art. (See, e.g., Rossi, J.J. (1995) Br. Med. Bull. 51(1):217-225; Boado, R.J. et al. (1998) J. Pharm. Sci. 87(11):1308-1315; and Morris, M.C. et al. (1997) Nucleic Acids Res. 25(14):2730-2736.)

In another embodiment of the invention, polynucleotides encoding IGSFP may be used for somatic or germline gene therapy. Gene therapy may be performed to (i) correct a genetic deficiency (e.g., in the cases of severe combined immunodeficiency (SCID)-X1 disease characterized by X-linked inheritance (Cavazzana-Calvo, M. et al. (2000) Science 288:669-672), severe combined immunodeficiency syndrome associated with an inherited adenosine deaminase (ADA) deficiency (Blaese, R.M. et al. (1995) Science 270:475-480; Bordignon, C. et al. (1995) Science 270:470-475), cystic fibrosis (Zabner, J. et al. (1993) Cell 75:207-216; Crystal, R.G. et al. (1995) Hum. Gene Therapy 6:643-666; Crystal, R.G. et al. (1995) Hum. Gene Therapy 6:667-703), thalassamias, familial hypercholesterolemia, and hemophilia resulting from Factor VIII or Factor IX deficiencies (Crystal, R.G. (1995) Science 270:404-410; Verma, I.M. and N. Somia (1997) Nature 389:239-242)), (ii) express a conditionally lethal gene product (e.g., in the case of cancers which result from unregulated cell proliferation), or (iii) express a protein which affords protection against intracellular parasites (e.g., against human retroviruses, such as human immunodeficiency virus (HIV) (Baltimore, D. (1988) Nature 335:395-396; Poeschla, E. et al. (1996) Proc. Natl. Acad. Sci. USA 93:11395-11399), hepatitis B or C virus (HBV, HCV); fungal parasites, such as Candida albicans and Paracoccidioides brasiliensis; and protozoan parasites such as Plasmodium falciparum and Trypanosoma cruzi). In the case where a genetic deficiency in IGSFP expression or regulation causes disease, the expression of IGSFP from an appropriate population of transduced cells may alleviate the clinical manifestations caused by the genetic deficiency.

In a further embodiment of the invention, diseases or disorders caused by deficiencies in IGSFP are treated by constructing mammalian expression vectors encoding IGSFP and introducing these vectors by mechanical means into IGSFP-deficient cells. Mechanical transfer technologies for use with cells in vivo or ex vitro include (i) direct DNA microinjection into individual cells, (ii) ballistic gold particle delivery, (iii) liposome-mediated transfection, (iv) receptor-mediated gene transfer, and (v) the use of DNA transposons (Morgan, R.A. and W.F. Anderson (1993) Annu. Rev. Biochem. 62:191-217; Ivics, Z. (1997) Cell 91:501-510; Boulay, J-L. and H. Récipon (1998) Curr. Opin. Biotechnol. 9:445-450).

Expression vectors that may be effective for the expression of IGSFP include, but are not limited to, the PCDNA 3.1, EPITAG, PRCCMV2, PREP, PVAX, PCR2-TOPOTA vectors

(Invitrogen, Carlsbad CA), PCMV-SCRIPT, PCMV-TAG, PEGSH/PERV (Stratagene, La Jolla CA), and PTET-OFF, PTET-ON, PTRE2, PTRE2-LUC, PTK-HYG (Clontech, Palo Alto CA). IGSFP may be expressed using (i) a constitutively active promoter, (e.g., from cytomegalovirus (CMV), Rous sarcoma virus (RSV), SV40 virus, thymidine kinase (TK), or β-actin genes), (ii) an inducible promoter (e.g., the tetracycline-regulated promoter (Gossen, M. and H. Bujard (1992) Proc. Natl. Acad. Sci. USA 89:5547-5551; Gossen, M. et al. (1995) Science 268:1766-1769; Rossi, F.M.V. and H.M. Blau (1998) Curr. Opin. Biotechnol. 9:451-456), commercially available in the T-REX plasmid (Invitrogen)); the ecdysone-inducible promoter (available in the plasmids PVGRXR and PIND; Invitrogen); the FK506/rapamycin inducible promoter; or the RU486/mifepristone inducible promoter (Rossi, F.M.V. and H.M. Blau, supra)), or (iii) a tissue-specific promoter or the native promoter of the endogenous gene encoding IGSFP from a normal individual.

Commercially available liposome transformation kits (e.g., the PERFECT LIPID TRANSFECTION KIT, available from Invitrogen) allow one with ordinary skill in the art to deliver polynucleotides to target cells in culture and require minimal effort to optimize experimental parameters. In the alternative, transformation is performed using the calcium phosphate method (Graham, F.L. and A.J. Eb (1973) Virology 52:456-467), or by electroporation (Neumann, E. et al. (1982) EMBO J. 1:841-845). The introduction of DNA to primary cells requires modification of these standardized mammalian transfection protocols.

In another embodiment of the invention, diseases or disorders caused by genetic defects with respect to IGSFP expression are treated by constructing a retrovirus vector consisting of (i) the polynucleotide encoding IGSFP under the control of an independent promoter or the retrovirus long terminal repeat (LTR) promoter, (ii) appropriate RNA packaging signals, and (iii) a Rev-responsive element (RRE) along with additional retrovirus *cis*-acting RNA sequences and coding sequences required for efficient vector propagation. Retrovirus vectors (e.g., PFB and PFBNEO) are commercially available (Stratagene) and are based on published data (Riviere, I. et al. (1995) Proc. Natl. Acad. Sci. USA 92:6733-6737), incorporated by reference herein. The vector is propagated in an appropriate vector producing cell line (VPCL) that expresses an envelope gene with a tropism for receptors on the target cells or a promiscuous envelope protein such as VSVg (Armentano, D. et al. (1987) J. Virol. 61:1647-1650; Bender, M.A. et al. (1987) J. Virol. 61:1639-1646; Adam, M.A. and A.D. Miller (1988) J. Virol. 62:3802-3806; Dull, T. et al. (1998) J. Virol. 72:8463-8471; Zufferey, R. et al. (1998) J. Virol. 72:9873-9880). U.S. Patent No. 5,910,434 to Rigg ("Method for obtaining retrovirus packaging cell lines producing high transducing efficiency retroviral supernatant") discloses a method for obtaining retrovirus packaging cell lines and is hereby incorporated by reference.

Propagation of retrovirus vectors, transduction of a population of cells (e.g., CD4+ T-cells), and the return of transduced cells to a patient are procedures well known to persons skilled in the art of gene therapy and have been well documented (Ranga, U. et al. (1997) J. Virol. 71:7020-7029; Bauer, G. et al. (1997) Blood 89:2259-2267; Bonyhadi, M.L. (1997) J. Virol. 71:4707-4716; Ranga, U. et al. (1998) Proc. Natl. Acad. Sci. USA 95:1201-1206; Su, L. (1997) Blood 89:2283-2290).

In the alternative, an adenovirus-based gene therapy delivery system is used to deliver polynucleotides encoding IGSFP to cells which have one or more genetic abnormalities with respect to the expression of IGSFP. The construction and packaging of adenovirus-based vectors are well known to those with ordinary skill in the art. Replication defective adenovirus vectors have proven to be versatile for importing genes encoding immunoregulatory proteins into intact islets in the pancreas (Csete, M.E. et al. (1995) Transplantation 27:263-268). Potentially useful adenoviral vectors are described in U.S. Patent No. 5,707,618 to Armentano ("Adenovirus vectors for gene therapy"), hereby incorporated by reference. For adenoviral vectors, see also Antinozzi, P.A. et al. (1999) Annu. Rev. Nutr. 19:511-544 and Verma, I.M. and N. Somia (1997) Nature 18:389:239-242, both incorporated by reference herein.

In another alternative, a herpes-based, gene therapy delivery system is used to deliver polynucleotides encoding IGSFP to target cells which have one or more genetic abnormalities with respect to the expression of IGSFP. The use of herpes simplex virus (HSV)-based vectors may be especially valuable for introducing IGSFP to cells of the central nervous system, for which HSV has a tropism. The construction and packaging of herpes-based vectors are well known to those with ordinary skill in the art. A replication-competent herpes simplex virus (HSV) type 1-based vector has been used to deliver a reporter gene to the eyes of primates (Liu, X. et al. (1999) Exp. Eye Res. 169:385-395). The construction of a HSV-1 virus vector has also been disclosed in detail in U.S. Patent No. 5,804,413 to DeLuca ("Herpes simplex virus strains for gene transfer"), which is hereby incorporated by reference. U.S. Patent No. 5,804,413 teaches the use of recombinant HSV d92 which consists of a genome containing at least one exogenous gene to be transferred to a cell under the control of the appropriate promoter for purposes including human gene therapy. Also taught by this patent are the construction and use of recombinant HSV strains deleted for ICP4, ICP27 and ICP22. For HSV vectors, see also Goins, W.F. et al. (1999) J. Virol. 73:519-532 and Xu, H. et al. (1994) Dev. Biol. 163:152-161, hereby incorporated by reference. The manipulation of cloned herpesvirus sequences, the generation of recombinant virus following the transfection of multiple plasmids containing different segments of the large herpesvirus genomes, the growth and propagation of herpesvirus, and the infection of cells with herpesvirus are techniques well known to those of

ordinary skill in the art.

In another alternative, an alphavirus (positive, single-stranded RNA virus) vector is used to deliver polynucleotides encoding IGSFP to target cells. The biology of the prototypic alphavirus, Semliki Forest Virus (SFV), has been studied extensively and gene transfer vectors have been based on the SFV genome (Garoff, H. and K.-J. Li (1998) Curr. Opin. Biotechnol. 9:464-469). During alphavirus RNA replication, a subgenomic RNA is generated that normally encodes the viral capsid proteins. This subgenomic RNA replicates to higher levels than the full length genomic RNA, resulting in the overproduction of capsid proteins relative to the viral proteins with enzymatic activity (e.g., protease and polymerase). Similarly, inserting the coding sequence for IGSFP into the alphavirus genome in place of the capsid-coding region results in the production of a large number of IGSFP-coding RNAs and the synthesis of high levels of IGSFP in vector transduced cells. While alphavirus infection is typically associated with cell lysis within a few days, the ability to establish a persistent infection in hamster normal kidney cells (BHK-21) with a variant of Sindbis virus (SIN) indicates that the lytic replication of alphaviruses can be altered to suit the needs of the gene therapy application (Dryga, S.A. et al. (1997) Virology 228:74-83). The wide host range of alphaviruses will allow the introduction of IGSFP into a variety of cell types. The specific transduction of a subset of cells in a population may require the sorting of cells prior to transduction. The methods of manipulating infectious cDNA clones of alphaviruses, performing alphavirus cDNA and RNA transfections, and performing alphavirus infections, are well known to those with ordinary skill in the art.

Oligonucleotides derived from the transcription initiation site, e.g., between about positions -10 and +10 from the start site, may also be employed to inhibit gene expression. Similarly, inhibition can be achieved using triple helix base-pairing methodology. Triple helix pairing is useful because it causes inhibition of the ability of the double helix to open sufficiently for the binding of polymerases, transcription factors, or regulatory molecules. Recent therapeutic advances using triplex DNA have been described in the literature. (See, e.g., Gee, J.E. et al. (1994) in Huber, B.E. and B.I. Carr, Molecular and Immunologic Approaches, Futura Publishing, Mt. Kisco NY, pp. 163-177.) A complementary sequence or antisense molecule may also be designed to block translation of mRNA by preventing the transcript from binding to ribosomes.

Ribozymes, enzymatic RNA molecules, may also be used to catalyze the specific cleavage of RNA. The mechanism of ribozyme action involves sequence-specific hybridization of the ribozyme molecule to complementary target RNA, followed by endonucleolytic cleavage. For example, engineered hammerhead motif ribozyme molecules may specifically and efficiently catalyze

endonucleolytic cleavage of sequences encoding IGSFP.

Specific ribozyme cleavage sites within any potential RNA target are initially identified by scanning the target molecule for ribozyme cleavage sites, including the following sequences: GUA, GUU, and GUC. Once identified, short RNA sequences of between 15 and 20 ribonucleotides, corresponding to the region of the target gene containing the cleavage site, may be evaluated for secondary structural features which may render the oligonucleotide inoperable. The suitability of candidate targets may also be evaluated by testing accessibility to hybridization with complementary oligonucleotides using ribonuclease protection assays.

Complementary ribonucleic acid molecules and ribozymes of the invention may be prepared by any method known in the art for the synthesis of nucleic acid molecules. These include techniques for chemically synthesizing oligonucleotides such as solid phase phosphoramidite chemical synthesis. Alternatively, RNA molecules may be generated by in vitro and in vivo transcription of DNA sequences encoding IGSFP. Such DNA sequences may be incorporated into a wide variety of vectors with suitable RNA polymerase promoters such as T7 or SP6. Alternatively, these cDNA constructs that synthesize complementary RNA, constitutively or inducibly, can be introduced into cell lines, cells, or tissues.

RNA molecules may be modified to increase intracellular stability and half-life. Possible modifications include, but are not limited to, the addition of flanking sequences at the 5' and/or 3' ends of the molecule, or the use of phosphorothioate or 2' O-methyl rather than phosphodiesterase linkages within the backbone of the molecule. This concept is inherent in the production of PNAs and can be extended in all of these molecules by the inclusion of nontraditional bases such as inosine, queosine, and wybutosine, as well as acetyl-, methyl-, thio-, and similarly modified forms of adenine, cytidine, guanine, thymine, and uridine which are not as easily recognized by endogenous endonucleases.

An additional embodiment of the invention encompasses a method for screening for a compound which is effective in altering expression of a polynucleotide encoding IGSFP. Compounds which may be effective in altering expression of a specific polynucleotide may include, but are not limited to, oligonucleotides, antisense oligonucleotides, triple helix-forming oligonucleotides, transcription factors and other polypeptide transcriptional regulators, and non-macromolecular chemical entities which are capable of interacting with specific polynucleotide sequences. Effective compounds may alter polynucleotide expression by acting as either inhibitors or promoters of polynucleotide expression. Thus, in the treatment of disorders associated with increased IGSFP expression or activity, a compound which specifically inhibits expression of the polynucleotide encoding IGSFP may be therapeutically useful, and in the treatment of disorders associated with

decreased IGSFP expression or activity, a compound which specifically promotes expression of the polynucleotide encoding IGSFP may be therapeutically useful.

At least one, and up to a plurality, of test compounds may be screened for effectiveness in altering expression of a specific polynucleotide. A test compound may be obtained by any method commonly known in the art, including chemical modification of a compound known to be effective in altering polynucleotide expression; selection from an existing, commercially-available or proprietary library of naturally-occurring or non-natural chemical compounds; rational design of a compound based on chemical and/or structural properties of the target polynucleotide; and selection from a library of chemical compounds created combinatorially or randomly. A sample comprising a polynucleotide encoding IGSFP is exposed to at least one test compound thus obtained. The sample may comprise, for example, an intact or permeabilized cell, or an in vitro cell-free or reconstituted biochemical system. Alterations in the expression of a polynucleotide encoding IGSFP are assayed by any method commonly known in the art. Typically, the expression of a specific nucleotide is detected by hybridization with a probe having a nucleotide sequence complementary to the sequence of the polynucleotide encoding IGSFP. The amount of hybridization may be quantified, thus forming the basis for a comparison of the expression of the polynucleotide both with and without exposure to one or more test compounds. Detection of a change in the expression of a polynucleotide exposed to a test compound indicates that the test compound is effective in altering the expression of the polynucleotide. A screen for a compound effective in altering expression of a specific polynucleotide can be carried out, for example, using a Schizosaccharomyces pombe gene expression system (Atkins, D. et al. (1999) U.S. Patent No. 5,932,435; Arndt, G.M. et al. (2000) Nucleic Acids Res. 28:E15) or a human cell line such as HeLa cell (Clarke, M.L. et al. (2000) Biochem. Biophys. Res. Commun. 268:8-13). A particular embodiment of the present invention involves screening a combinatorial library of oligonucleotides (such as deoxyribonucleotides, ribonucleotides, peptide nucleic acids, and modified oligonucleotides) for antisense activity against a specific polynucleotide sequence (Bruice, T.W. et al. (1997) U.S. Patent No. 5,686,242; Bruice, T.W. et al. (2000) U.S. Patent No. 6,022,691).

Many methods for introducing vectors into cells or tissues are available and equally suitable for use in vivo, in vitro, and ex vivo. For ex vivo therapy, vectors may be introduced into stem cells taken from the patient and clonally propagated for autologous transplant back into that same patient. Delivery by transfection, by liposome injections, or by polycationic amino polymers may be achieved using methods which are well known in the art. (See, e.g., Goldman, C.K. et al. (1997) Nat. Biotechnol. 15:462-466.)

Any of the therapeutic methods described above may be applied to any subject in need of

such therapy, including, for example, mammals such as humans, dogs, cats, cows, horses, rabbits, and monkeys.

An additional embodiment of the invention relates to the administration of a composition which generally comprises an active ingredient formulated with a pharmaceutically acceptable excipient. Excipients may include, for example, sugars, starches, celluloses, gums, and proteins. Various formulations are commonly known and are thoroughly discussed in the latest edition of Remington's Pharmaceutical Sciences (Maack Publishing, Easton PA). Such compositions may consist of IGSFP, antibodies to IGSFP, and mimetics, agonists, antagonists, or inhibitors of IGSFP.

The compositions utilized in this invention may be administered by any number of routes including, but not limited to, oral, intravenous, intramuscular, intra-arterial, intramedullary, intrathecal, intraventricular, pulmonary, transdermal, subcutaneous, intraperitoneal, intranasal, enteral, topical, sublingual, or rectal means.

Compositions for pulmonary administration may be prepared in liquid or dry powder form. These compositions are generally aerosolized immediately prior to inhalation by the patient. In the case of small molecules (e.g. traditional low molecular weight organic drugs), aerosol delivery of fast-acting formulations is well-known in the art. In the case of macromolecules (e.g. larger peptides and proteins), recent developments in the field of pulmonary delivery via the alveolar region of the lung have enabled the practical delivery of drugs such as insulin to blood circulation (see, e.g., Patton, J.S. et al., U.S. Patent No. 5,997,848). Pulmonary delivery has the advantage of administration without needle injection, and obviates the need for potentially toxic penetration enhancers.

Compositions suitable for use in the invention include compositions wherein the active ingredients are contained in an effective amount to achieve the intended purpose. The determination of an effective dose is well within the capability of those skilled in the art.

Specialized forms of compositions may be prepared for direct intracellular delivery of macromolecules comprising IGSFP or fragments thereof. For example, liposome preparations containing a cell-impermeable macromolecule may promote cell fusion and intracellular delivery of the macromolecule. Alternatively, IGSFP or a fragment thereof may be joined to a short cationic N-terminal portion from the HIV Tat-1 protein. Fusion proteins thus generated have been found to transduce into the cells of all tissues, including the brain, in a mouse model system (Schwarze, S.R. et al. (1999) Science 285:1569-1572).

For any compound, the therapeutically effective dose can be estimated initially either in cell culture assays, e.g., of neoplastic cells, or in animal models such as mice, rats, rabbits, dogs, monkeys, or pigs. An animal model may also be used to determine the appropriate concentration range and

route of administration. Such information can then be used to determine useful doses and routes for administration in humans.

A therapeutically effective dose refers to that amount of active ingredient, for example IGSFP or fragments thereof, antibodies of IGSFP, and agonists, antagonists or inhibitors of IGSFP, which ameliorates the symptoms or condition. Therapeutic efficacy and toxicity may be determined by standard pharmaceutical procedures in cell cultures or with experimental animals, such as by calculating the ED_{50} (the dose therapeutically effective in 50% of the population) or LD_{50} (the dose lethal to 50% of the population) statistics. The dose ratio of toxic to therapeutic effects is the therapeutic index, which can be expressed as the LD_{50}/ED_{50} ratio. Compositions which exhibit large therapeutic indices are preferred. The data obtained from cell culture assays and animal studies are used to formulate a range of dosage for human use. The dosage contained in such compositions is preferably within a range of circulating concentrations that includes the ED_{50} with little or no toxicity. The dosage varies within this range depending upon the dosage form employed, the sensitivity of the patient, and the route of administration.

The exact dosage will be determined by the practitioner, in light of factors related to the subject requiring treatment. Dosage and administration are adjusted to provide sufficient levels of the active moiety or to maintain the desired effect. Factors which may be taken into account include the severity of the disease state, the general health of the subject, the age, weight, and gender of the subject, time and frequency of administration, drug combination(s), reaction sensitivities, and response to therapy. Long-acting compositions may be administered every 3 to 4 days, every week, or biweekly depending on the half-life and clearance rate of the particular formulation.

Normal dosage amounts may vary from about 0.1 μg to 100,000 μg, up to a total dose of about 1 gram, depending upon the route of administration. Guidance as to particular dosages and methods of delivery is provided in the literature and generally available to practitioners in the art. Those skilled in the art will employ different formulations for nucleotides than for proteins or their inhibitors. Similarly, delivery of polynucleotides or polypeptides will be specific to particular cells, conditions, locations, etc.

DIAGNOSTICS

In another embodiment, antibodies which specifically bind IGSFP may be used for the diagnosis of disorders characterized by expression of IGSFP, or in assays to monitor patients being treated with IGSFP or agonists, antagonists, or inhibitors of IGSFP. Antibodies useful for diagnostic purposes may be prepared in the same manner as described above for therapeutics. Diagnostic assays for IGSFP include methods which utilize the antibody and a label to detect IGSFP in human

body fluids or in extracts of cells or tissues. The antibodies may be used with or without modification, and may be labeled by covalent or non-covalent attachment of a reporter molecule. A wide variety of reporter molecules, several of which are described above, are known in the art and may be used.

A variety of protocols for measuring IGSFP, including ELISAs, RIAs, and FACS, are known in the art and provide a basis for diagnosing altered or abnormal levels of IGSFP expression. Normal or standard values for IGSFP expression are established by combining body fluids or cell extracts taken from normal mammalian subjects, for example, human subjects, with antibodies to IGSFP under conditions suitable for complex formation. The amount of standard complex formation may be quantitated by various methods, such as photometric means. Quantities of IGSFP expressed in subject, control, and disease samples from biopsied tissues are compared with the standard values. Deviation between standard and subject values establishes the parameters for diagnosing disease.

In another embodiment of the invention, the polynucleotides encoding IGSFP may be used for diagnostic purposes. The polynucleotides which may be used include oligonucleotide sequences, complementary RNA and DNA molecules, and PNAs. The polynucleotides may be used to detect and quantify gene expression in biopsied tissues in which expression of IGSFP may be correlated with disease. The diagnostic assay may be used to determine absence, presence, and excess expression of IGSFP, and to monitor regulation of IGSFP levels during therapeutic intervention.

In one aspect, hybridization with PCR probes which are capable of detecting polynucleotide sequences, including genomic sequences, encoding IGSFP or closely related molecules may be used to identify nucleic acid sequences which encode IGSFP. The specificity of the probe, whether it is made from a highly specific region, e.g., the 5' regulatory region, or from a less specific region, e.g., a conserved motif, and the stringency of the hybridization or amplification will determine whether the probe identifies only naturally occurring sequences encoding IGSFP, allelic variants, or related sequences.

Probes may also be used for the detection of related sequences, and may have at least 50% sequence identity to any of the IGSFP encoding sequences. The hybridization probes of the subject invention may be DNA or RNA and may be derived from the sequence of SEQ ID NO:13-24 or from genomic sequences including promoters, enhancers, and introns of the IGSFP gene.

Means for producing specific hybridization probes for DNAs encoding IGSFP include the cloning of polynucleotide sequences encoding IGSFP or IGSFP derivatives into vectors for the production of mRNA probes. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes in vitro by means of the addition of the appropriate RNA polymerases and the appropriate labeled nucleotides. Hybridization probes may be labeled by a

variety of reporter groups, for example, by radionuclides such as ^{32}P or ^{35}S, or by enzymatic labels, such as alkaline phosphatase coupled to the probe via avidin/biotin coupling systems, and the like.

Polynucleotide sequences encoding IGSFP may be used for the diagnosis of disorders associated with expression of IGSFP. Examples of such disorders include, but are not limited to, an immune system disorder such as acquired immunodeficiency syndrome (AIDS), X-linked agammaglobinemia of Bruton, common variable immunodeficiency (CVI), DiGeorge's syndrome (thymic hypoplasia), thymic dysplasia, isolated IgA deficiency, severe combined immunodeficiency disease (SCID), immunodeficiency with thrombocytopenia and eczema (Wiskott-Aldrich syndrome), Chediak-Higashi syndrome, chronic granulomatous diseases, hereditary angioneurotic edema, immunodeficiency associated with Cushing's disease, Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis, autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED), bronchitis, cholecystitis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, episodic lymphopenia with lymphocytotoxins, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, hypereosinophilia, irritable bowel syndrome, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, viral, bacterial, fungal, parasitic, protozoal, and helminthic infections, and trauma; a neurological disorder such as epilepsy, ischemic cerebrovascular disease, stroke, cerebral neoplasms, Alzheimer's disease, Pick's disease, Huntington's disease, dementia, Parkinson's disease and other extrapyramidal disorders, amyotrophic lateral sclerosis and other motor neuron disorders, progressive neural muscular atrophy, retinitis pigmentosa, hereditary ataxias, multiple sclerosis and other demyelinating diseases, bacterial and viral meningitis, brain abscess, subdural empyema, epidural abscess, suppurative intracranial thrombophlebitis, myelitis and radiculitis, viral central nervous system disease, prion diseases including kuru, Creutzfeldt-Jakob disease, and Gerstmann-Straussler-Scheinker syndrome, fatal familial insomnia, nutritional and metabolic diseases of the nervous system, neurofibromatosis, tuberous sclerosis, cerebelloretinal hemangioblastomatosis, encephalotrigeminal syndrome, mental retardation and other developmental disorders of the central nervous system including Down syndrome, cerebral palsy, neuroskeletal disorders, autonomic nervous system disorders, cranial nerve disorders, spinal cord diseases, muscular dystrophy and other

neuromuscular disorders, peripheral nervous system disorders, dermatomyositis and polymyositis, inherited, metabolic, endocrine, and toxic myopathies, myasthenia gravis, periodic paralysis, mental disorders including mood, anxiety, and schizophrenic disorders, seasonal affective disorder (SAD), akathesia, amnesia, catatonia, diabetic neuropathy, tardive dyskinesia, dystonias, paranoid psychoses, postherpetic neuralgia, Tourette's disorder, progressive supranuclear palsy, corticobasal degeneration, and familial frontotemporal dementia; a developmental disorder such as renal tubular acidosis, anemia, Cushing's syndrome, achondroplastic dwarfism, Duchenne and Becker muscular dystrophy, epilepsy, gonadal dysgenesis, WAGR syndrome (Wilms' tumor, aniridia, genitourinary abnormalities, and mental retardation), Smith-Magenis syndrome, myelodysplastic syndrome, hereditary mucoepithelial dysplasia, hereditary keratodermas, hereditary neuropathies such as Charcot-Marie-Tooth disease and neurofibromatosis, hypothyroidism, hydrocephalus, seizure disorders such as Syndenham's chorea and cerebral palsy, spina bifida, anencephaly, craniorachischisis, congenital glaucoma, cataract, and sensorineural hearing loss; a muscle disorder such as cardiomyopathy, myocarditis, Duchenne's muscular dystrophy, Becker's muscular dystrophy, myotonic dystrophy, central core disease, nemaline myopathy, centronuclear myopathy, lipid myopathy, mitochondrial myopathy, infectious myositis, polymyositis, dermatomyositis, inclusion body myositis, thyrotoxic myopathy, and ethanol myopathy; and a cell proliferative disorder such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, and cancers including adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus. The polynucleotide sequences encoding IGSFP may be used in Southern or northern analysis, dot blot, or other membrane-based technologies; in PCR technologies; in dipstick, pin, and multiformat ELISA-like assays; and in microarrays utilizing fluids or tissues from patients to detect altered IGSFP expression. Such qualitative or quantitative methods are well known in the art.

In a particular aspect, the nucleotide sequences encoding IGSFP may be useful in assays that detect the presence of associated disorders, particularly those mentioned above. The nucleotide sequences encoding IGSFP may be labeled by standard methods and added to a fluid or tissue sample from a patient under conditions suitable for the formation of hybridization complexes. After a suitable incubation period, the sample is washed and the signal is quantified and compared with a standard value. If the amount of signal in the patient sample is significantly altered in comparison to a control

sample then the presence of altered levels of nucleotide sequences encoding IGSFP in the sample indicates the presence of the associated disorder. Such assays may also be used to evaluate the efficacy of a particular therapeutic treatment regimen in animal studies, in clinical trials, or to monitor the treatment of an individual patient.

In order to provide a basis for the diagnosis of a disorder associated with expression of IGSFP, a normal or standard profile for expression is established. This may be accomplished by combining body fluids or cell extracts taken from normal subjects, either animal or human, with a sequence, or a fragment thereof, encoding IGSFP, under conditions suitable for hybridization or amplification. Standard hybridization may be quantified by comparing the values obtained from normal subjects with values from an experiment in which a known amount of a substantially purified polynucleotide is used. Standard values obtained in this manner may be compared with values obtained from samples from patients who are symptomatic for a disorder. Deviation from standard values is used to establish the presence of a disorder.

Once the presence of a disorder is established and a treatment protocol is initiated, hybridization assays may be repeated on a regular basis to determine if the level of expression in the patient begins to approximate that which is observed in the normal subject. The results obtained from successive assays may be used to show the efficacy of treatment over a period ranging from several days to months.

With respect to cancer, the presence of an abnormal amount of transcript (either under- or overexpressed) in biopsied tissue from an individual may indicate a predisposition for the development of the disease, or may provide a means for detecting the disease prior to the appearance of actual clinical symptoms. A more definitive diagnosis of this type may allow health professionals to employ preventative measures or aggressive treatment earlier thereby preventing the development or further progression of the cancer.

Additional diagnostic uses for oligonucleotides designed from the sequences encoding IGSFP may involve the use of PCR. These oligomers may be chemically synthesized, generated enzymatically, or produced in vitro. Oligomers will preferably contain a fragment of a polynucleotide encoding IGSFP, or a fragment of a polynucleotide complementary to the polynucleotide encoding IGSFP, and will be employed under optimized conditions for identification of a specific gene or condition. Oligomers may also be employed under less stringent conditions for detection or quantification of closely related DNA or RNA sequences.

In a particular aspect, oligonucleotide primers derived from the polynucleotide sequences encoding IGSFP may be used to detect single nucleotide polymorphisms (SNPs). SNPs are

substitutions, insertions and deletions that are a frequent cause of inherited or acquired genetic disease in humans. Methods of SNP detection include, but are not limited to, single-stranded conformation polymorphism (SSCP) and fluorescent SSCP (fSSCP) methods. In SSCP, oligonucleotide primers derived from the polynucleotide sequences encoding IGSFP are used to amplify DNA using the polymerase chain reaction (PCR). The DNA may be derived, for example, from diseased or normal tissue, biopsy samples, bodily fluids, and the like. SNPs in the DNA cause differences in the secondary and tertiary structures of PCR products in single-stranded form, and these differences are detectable using gel electrophoresis in non-denaturing gels. In fSCCP, the oligonucleotide primers are fluorescently labeled, which allows detection of the amplimers in high-throughput equipment such as DNA sequencing machines. Additionally, sequence database analysis methods, termed in silico SNP (isSNP), are capable of identifying polymorphisms by comparing the sequence of individual overlapping DNA fragments which assemble into a common consensus sequence. These computer-based methods filter out sequence variations due to laboratory preparation of DNA and sequencing errors using statistical models and automated analyses of DNA sequence chromatograms. In the alternative, SNPs may be detected and characterized by mass spectrometry using, for example, the high throughput MASSARRAY system (Sequenom, Inc., San Diego CA).

SNPs may be used to study the genetic basis of human disease. For example, at least 16 common SNPs have been associated with non-insulin-dependent diabetes mellitus. SNPs are also useful for examining differences in disease outcomes in monogenic disorders, such as cystic fibrosis, sickle cell anemia, or chronic granulomatous disease. For example, variants in the mannose-binding lectin, MBL2, have been shown to be correlated with deleterious pulmonary outcomes in cystic fibrosis. SNPs also have utility in pharmacogenomics, the identification of genetic variants that influence a patient's response to a drug, such as life-threatening toxicity. For example, a variation in N-acetyl transferase is associated with a high incidence of peripheral neuropathy in response to the anti-tuberculosis drug isoniazid, while a variation in the core promoter of the ALOX5 gene results in diminished clinical response to treatment with an anti-asthma drug that targets the 5-lipoxygenase pathway. Analysis of the distribution of SNPs in different populations is useful for investigating genetic drift, mutation, recombination, and selection, as well as for tracing the origins of populations and their migrations. (Taylor, J.G. et al. (2001) Trends Mol. Med. 7:507-512; Kwok, P.-Y. and Z. Gu (1999) Mol. Med. Today 5:538-543; Nowotny, P. et al. (2001) Curr. Opin. Neurobiol. 11:637-641.)

Methods which may also be used to quantify the expression of IGSFP include radiolabeling or biotinylating nucleotides, coamplification of a control nucleic acid, and interpolating results from standard curves. (See, e.g., Melby, P.C. et al. (1993) J. Immunol. Methods 159:235-244; Duplaa, C.

et al. (1993) Anal. Biochem. 212:229-236.) The speed of quantitation of multiple samples may be accelerated by running the assay in a high-throughput format where the oligomer or polynucleotide of interest is presented in various dilutions and a spectrophotometric or colorimetric response gives rapid quantitation.

In further embodiments, oligonucleotides or longer fragments derived from any of the polynucleotide sequences described herein may be used as elements on a microarray. The microarray can be used in transcript imaging techniques which monitor the relative expression levels of large numbers of genes simultaneously as described below. The microarray may also be used to identify genetic variants, mutations, and polymorphisms. This information may be used to determine gene function, to understand the genetic basis of a disorder, to diagnose a disorder, to monitor progression/regression of disease as a function of gene expression, and to develop and monitor the activities of therapeutic agents in the treatment of disease. In particular, this information may be used to develop a pharmacogenomic profile of a patient in order to select the most appropriate and effective treatment regimen for that patient. For example, therapeutic agents which are highly effective and display the fewest side effects may be selected for a patient based on his/her pharmacogenomic profile.

In another embodiment, IGSFP, fragments of IGSFP, or antibodies specific for IGSFP may be used as elements on a microarray. The microarray may be used to monitor or measure protein-protein interactions, drug-target interactions, and gene expression profiles, as described above.

A particular embodiment relates to the use of the polynucleotides of the present invention to generate a transcript image of a tissue or cell type. A transcript image represents the global pattern of gene expression by a particular tissue or cell type. Global gene expression patterns are analyzed by quantifying the number of expressed genes and their relative abundance under given conditions and at a given time. (See Seilhamer et al., "Comparative Gene Transcript Analysis," U.S. Patent No. 5,840,484, expressly incorporated by reference herein.) Thus a transcript image may be generated by hybridizing the polynucleotides of the present invention or their complements to the totality of transcripts or reverse transcripts of a particular tissue or cell type. In one embodiment, the hybridization takes place in high-throughput format, wherein the polynucleotides of the present invention or their complements comprise a subset of a plurality of elements on a microarray. The resultant transcript image would provide a profile of gene activity.

Transcript images may be generated using transcripts isolated from tissues, cell lines, biopsies, or other biological samples. The transcript image may thus reflect gene expression in vivo, as in the case of a tissue or biopsy sample, or in vitro, as in the case of a cell line.

Transcript images which profile the expression of the polynucleotides of the present invention may also be used in conjunction with in vitro model systems and preclinical evaluation of pharmaceuticals, as well as toxicological testing of industrial and naturally-occurring environmental compounds. All compounds induce characteristic gene expression patterns, frequently termed molecular fingerprints or toxicant signatures, which are indicative of mechanisms of action and toxicity (Nuwaysir, E.F. et al. (1999) Mol. Carcinog. 24:153-159; Steiner, S. and N.L. Anderson (2000) Toxicol. Lett. 112-113:467-471, expressly incorporated by reference herein). If a test compound has a signature similar to that of a compound with known toxicity, it is likely to share those toxic properties. These fingerprints or signatures are most useful and refined when they contain expression information from a large number of genes and gene families. Ideally, a genome-wide measurement of expression provides the highest quality signature. Even genes whose expression is not altered by any tested compounds are important as well, as the levels of expression of these genes are used to normalize the rest of the expression data. The normalization procedure is useful for comparison of expression data after treatment with different compounds. While the assignment of gene function to elements of a toxicant signature aids in interpretation of toxicity mechanisms, knowledge of gene function is not necessary for the statistical matching of signatures which leads to prediction of toxicity. (See, for example, Press Release 00-02 from the National Institute of Environmental Health Sciences, released February 29, 2000, available at http://www.niehs.nih.gov/oc/news/toxchip.htm.) Therefore, it is important and desirable in toxicological screening using toxicant signatures to include all expressed gene sequences.

In one embodiment, the toxicity of a test compound is assessed by treating a biological sample containing nucleic acids with the test compound. Nucleic acids that are expressed in the treated biological sample are hybridized with one or more probes specific to the polynucleotides of the present invention, so that transcript levels corresponding to the polynucleotides of the present invention may be quantified. The transcript levels in the treated biological sample are compared with levels in an untreated biological sample. Differences in the transcript levels between the two samples are indicative of a toxic response caused by the test compound in the treated sample.

Another particular embodiment relates to the use of the polypeptide sequences of the present invention to analyze the proteome of a tissue or cell type. The term proteome refers to the global pattern of protein expression in a particular tissue or cell type. Each protein component of a proteome can be subjected individually to further analysis. Proteome expression patterns, or profiles, are analyzed by quantifying the number of expressed proteins and their relative abundance under given conditions and at a given time. A profile of a cell's proteome may thus be generated by separating

and analyzing the polypeptides of a particular tissue or cell type. In one embodiment, the separation is achieved using two-dimensional gel electrophoresis, in which proteins from a sample are separated by isoelectric focusing in the first dimension, and then according to molecular weight by sodium dodecyl sulfate slab gel electrophoresis in the second dimension (Steiner and Anderson, supra). The proteins are visualized in the gel as discrete and uniquely positioned spots, typically by staining the gel with an agent such as Coomassie Blue or silver or fluorescent stains. The optical density of each protein spot is generally proportional to the level of the protein in the sample. The optical densities of equivalently positioned protein spots from different samples, for example, from biological samples either treated or untreated with a test compound or therapeutic agent, are compared to identify any changes in protein spot density related to the treatment. The proteins in the spots are partially sequenced using, for example, standard methods employing chemical or enzymatic cleavage followed by mass spectrometry. The identity of the protein in a spot may be determined by comparing its partial sequence, preferably of at least 5 contiguous amino acid residues, to the polypeptide sequences of the present invention. In some cases, further sequence data may be obtained for definitive protein identification.

A proteomic profile may also be generated using antibodies specific for IGSFP to quantify the levels of IGSFP expression. In one embodiment, the antibodies are used as elements on a microarray, and protein expression levels are quantified by exposing the microarray to the sample and detecting the levels of protein bound to each array element (Lueking, A. et al. (1999) Anal. Biochem. 270:103-111; Mendoze, L.G. et al. (1999) Biotechniques 27:778-788). Detection may be performed by a variety of methods known in the art, for example, by reacting the proteins in the sample with a thiol- or amino-reactive fluorescent compound and detecting the amount of fluorescence bound at each array element.

Toxicant signatures at the proteome level are also useful for toxicological screening, and should be analyzed in parallel with toxicant signatures at the transcript level. There is a poor correlation between transcript and protein abundances for some proteins in some tissues (Anderson, N.L. and J. Seilhamer (1997) Electrophoresis 18:533-537), so proteome toxicant signatures may be useful in the analysis of compounds which do not significantly affect the transcript image, but which alter the proteomic profile. In addition, the analysis of transcripts in body fluids is difficult, due to rapid degradation of mRNA, so proteomic profiling may be more reliable and informative in such cases.

In another embodiment, the toxicity of a test compound is assessed by treating a biological sample containing proteins with the test compound. Proteins that are expressed in the treated biological sample are separated so that the amount of each protein can be quantified. The amount of

each protein is compared to the amount of the corresponding protein in an untreated biological sample. A difference in the amount of protein between the two samples is indicative of a toxic response to the test compound in the treated sample. Individual proteins are identified by sequencing the amino acid residues of the individual proteins and comparing these partial sequences to the polypeptides of the present invention.

In another embodiment, the toxicity of a test compound is assessed by treating a biological sample containing proteins with the test compound. Proteins from the biological sample are incubated with antibodies specific to the polypeptides of the present invention. The amount of protein recognized by the antibodies is quantified. The amount of protein in the treated biological sample is compared with the amount in an untreated biological sample. A difference in the amount of protein between the two samples is indicative of a toxic response to the test compound in the treated sample.

Microarrays may be prepared, used, and analyzed using methods known in the art. (See, e.g., Brennan, T.M. et al. (1995) U.S. Patent No. 5,474,796; Schena, M. et al. (1996) Proc. Natl. Acad. Sci. USA 93:10614-10619; Baldeschweiler et al. (1995) PCT application WO95/251116; Shalon, D. et al. (1995) PCT application WO95/35505; Heller, R.A. et al. (1997) Proc. Natl. Acad. Sci. USA 94:2150-2155; and Heller, M.J. et al. (1997) U.S. Patent No. 5,605,662.) Various types of microarrays are well known and thoroughly described in DNA Microarrays: A Practical Approach, M. Schena, ed. (1999) Oxford University Press, London, hereby expressly incorporated by reference.

In another embodiment of the invention, nucleic acid sequences encoding IGSFP may be used to generate hybridization probes useful in mapping the naturally occurring genomic sequence. Either coding or noncoding sequences may be used, and in some instances, noncoding sequences may be preferable over coding sequences. For example, conservation of a coding sequence among members of a multi-gene family may potentially cause undesired cross hybridization during chromosomal mapping. The sequences may be mapped to a particular chromosome, to a specific region of a chromosome, or to artificial chromosome constructions, e.g., human artificial chromosomes (HACs), yeast artificial chromosomes (YACs), bacterial artificial chromosomes (BACs), bacterial P1 constructions, or single chromosome cDNA libraries. (See, e.g., Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355; Price, C.M. (1993) Blood Rev. 7:127-134; and Trask, B.J. (1991) Trends Genet. 7:149-154.) Once mapped, the nucleic acid sequences of the invention may be used to develop genetic linkage maps, for example, which correlate the inheritance of a disease state with the inheritance of a particular chromosome region or restriction fragment length polymorphism (RFLP). (See, for example, Lander, E.S. and D. Botstein (1986) Proc. Natl. Acad. Sci. USA 83:7353-7357.)

Fluorescent in situ hybridization (FISH) may be correlated with other physical and genetic

map data. (See, e.g., Heinz-Ulrich, et al. (1995) in Meyers, supra, pp. 965-968.) Examples of genetic map data can be found in various scientific journals or at the Online Mendelian Inheritance in Man (OMIM) World Wide Web site. Correlation between the location of the gene encoding IGSFP on a physical map and a specific disorder, or a predisposition to a specific disorder, may help define the region of DNA associated with that disorder and thus may further positional cloning efforts.

In situ hybridization of chromosomal preparations and physical mapping techniques, such as linkage analysis using established chromosomal markers, may be used for extending genetic maps. Often the placement of a gene on the chromosome of another mammalian species, such as mouse, may reveal associated markers even if the exact chromosomal locus is not known. This information is valuable to investigators searching for disease genes using positional cloning or other gene discovery techniques. Once the gene or genes responsible for a disease or syndrome have been crudely localized by genetic linkage to a particular genomic region, e.g., ataxia-telangiectasia to 11q22-23, any sequences mapping to that area may represent associated or regulatory genes for further investigation. (See, e.g., Gatti, R.A. et al. (1988) Nature 336:577-580.) The nucleotide sequence of the instant invention may also be used to detect differences in the chromosomal location due to translocation, inversion, etc., among normal, carrier, or affected individuals.

In another embodiment of the invention, IGSFP, its catalytic or immunogenic fragments, or oligopeptides thereof can be used for screening libraries of compounds in any of a variety of drug screening techniques. The fragment employed in such screening may be free in solution, affixed to a solid support, borne on a cell surface, or located intracellularly. The formation of binding complexes between IGSFP and the agent being tested may be measured.

Another technique for drug screening provides for high throughput screening of compounds having suitable binding affinity to the protein of interest. (See, e.g., Geysen, et al. (1984) PCT application WO84/03564.) In this method, large numbers of different small test compounds are synthesized on a solid substrate. The test compounds are reacted with IGSFP, or fragments thereof, and washed. Bound IGSFP is then detected by methods well known in the art. Purified IGSFP can also be coated directly onto plates for use in the aforementioned drug screening techniques. Alternatively, non-neutralizing antibodies can be used to capture the peptide and immobilize it on a solid support.

In another embodiment, one may use competitive drug screening assays in which neutralizing antibodies capable of binding IGSFP specifically compete with a test compound for binding IGSFP. In this manner, antibodies can be used to detect the presence of any peptide which shares one or more antigenic determinants with IGSFP.

In additional embodiments, the nucleotide sequences which encode IGSFP may be used in any molecular biology techniques that have yet to be developed, provided the new techniques rely on properties of nucleotide sequences that are currently known, including, but not limited to, such properties as the triplet genetic code and specific base pair interactions.

Without further elaboration, it is believed that one skilled in the art can, using the preceding description, utilize the present invention to its fullest extent. The following embodiments are, therefore, to be construed as merely illustrative, and not limitative of the remainder of the disclosure in any way whatsoever.

The disclosures of all patents, applications and publications, mentioned above and below, including U.S. Ser. No.60/275,249, U.S. Ser. No.60/316,810, U.S. Ser. No.60/323,977, U.S. Ser. No.60/348,447, and U.S. Ser. No.60/343,880, are expressly incorporated by reference herein.

EXAMPLES

I. Construction of cDNA Libraries

Incyte cDNAs were derived from cDNA libraries described in the LIFESEQ GOLD database (Incyte Genomics, Palo Alto CA). Some tissues were homogenized and lysed in guanidinium isothiocyanate, while others were homogenized and lysed in phenol or in a suitable mixture of denaturants, such as TRIZOL (Life Technologies), a monophasic solution of phenol and guanidine isothiocyanate. The resulting lysates were centrifuged over CsCl cushions or extracted with chloroform. RNA was precipitated from the lysates with either isopropanol or sodium acetate and ethanol, or by other routine methods.

Phenol extraction and precipitation of RNA were repeated as necessary to increase RNA purity. In some cases, RNA was treated with DNase. For most libraries, poly(A)+ RNA was isolated using oligo d(T)-coupled paramagnetic particles (Promega), OLIGOTEX latex particles (QIAGEN, Chatsworth CA), or an OLIGOTEX mRNA purification kit (QIAGEN). Alternatively, RNA was isolated directly from tissue lysates using other RNA isolation kits, e.g., the POLY(A)PURE mRNA purification kit (Ambion, Austin TX).

In some cases, Stratagene was provided with RNA and constructed the corresponding cDNA libraries. Otherwise, cDNA was synthesized and cDNA libraries were constructed with the UNIZAP vector system (Stratagene) or SUPERSCRIPT plasmid system (Life Technologies), using the recommended procedures or similar methods known in the art. (See, e.g., Ausubel, 1997, supra, units 5.1-6.6.) Reverse transcription was initiated using oligo d(T) or random primers. Synthetic oligonucleotide adapters were ligated to double stranded cDNA, and the cDNA was digested with the

appropriate restriction enzyme or enzymes. For most libraries, the cDNA was size-selected (300-1000 bp) using SEPHACRYL S1000, SEPHAROSE CL2B, or SEPHAROSE CL4B column chromatography (Amersham Pharmacia Biotech) or preparative agarose gel electrophoresis. cDNAs were ligated into compatible restriction enzyme sites of the polylinker of a suitable plasmid, e.g., PBLUESCRIPT plasmid (Stratagene), PSPORT1 plasmid (Life Technologies), PCDNA2.1 plasmid (Invitrogen, Carlsbad CA), PBK-CMV plasmid (Stratagene), PCR2-TOPOTA plasmid (Invitrogen), PCMV-ICIS plasmid (Stratagene), pIGEN (Incyte Genomics, Palo Alto CA), pRARE (Incyte Genomics), or pINCY (Incyte Genomics), or derivatives thereof. Recombinant plasmids were transformed into competent E. coli cells including XL1-Blue, XL1-BlueMRF, or SOLR from Stratagene or DH5α, DH10B, or ElectroMAX DH10B from Life Technologies.

II. Isolation of cDNA Clones

Plasmids obtained as described in Example I were recovered from host cells by in vivo excision using the UNIZAP vector system (Stratagene) or by cell lysis. Plasmids were purified using at least one of the following: a Magic or WIZARD Minipreps DNA purification system (Promega); an AGTC Miniprep purification kit (Edge Biosystems, Gaithersburg MD); and QIAWELL 8 Plasmid, QIAWELL 8 Plus Plasmid, QIAWELL 8 Ultra Plasmid purification systems or the R.E.A.L. PREP 96 plasmid purification kit from QIAGEN. Following precipitation, plasmids were resuspended in 0.1 ml of distilled water and stored, with or without lyophilization, at 4°C.

Alternatively, plasmid DNA was amplified from host cell lysates using direct link PCR in a high-throughput format (Rao, V.B. (1994) Anal. Biochem. 216:1-14). Host cell lysis and thermal cycling steps were carried out in a single reaction mixture. Samples were processed and stored in 384-well plates, and the concentration of amplified plasmid DNA was quantified fluorometrically using PICOGREEN dye (Molecular Probes, Eugene OR) and a FLUOROSKAN II fluorescence scanner (Labsystems Oy, Helsinki, Finland).

III. Sequencing and Analysis

Incyte cDNA recovered in plasmids as described in Example II were sequenced as follows. Sequencing reactions were processed using standard methods or high-throughput instrumentation such as the ABI CATALYST 800 (Applied Biosystems) thermal cycler or the PTC-200 thermal cycler (MJ Research) in conjunction with the HYDRA microdispenser (Robbins Scientific) or the MICROLAB 2200 (Hamilton) liquid transfer system. cDNA sequencing reactions were prepared using reagents provided by Amersham Pharmacia Biotech or supplied in ABI sequencing kits such as the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (Applied Biosystems). Electrophoretic separation of cDNA sequencing reactions and detection of labeled polynucleotides

were carried out using the MEGABACE 1000 DNA sequencing system (Molecular Dynamics); the ABI PRISM 373 or 377 sequencing system (Applied Biosystems) in conjunction with standard ABI protocols and base calling software; or other sequence analysis systems known in the art. Reading frames within the cDNA sequences were identified using standard methods (reviewed in Ausubel, 1997, supra, unit 7.7). Some of the cDNA sequences were selected for extension using the techniques disclosed in Example VIII.

The polynucleotide sequences derived from Incyte cDNAs were validated by removing vector, linker, and poly(A) sequences and by masking ambiguous bases, using algorithms and programs based on BLAST, dynamic programming, and dinucleotide nearest neighbor analysis. The Incyte cDNA sequences or translations thereof were then queried against a selection of public databases such as the GenBank primate, rodent, mammalian, vertebrate, and eukaryote databases, and BLOCKS, PRINTS, DOMO, PRODOM; PROTEOME databases with sequences from Homo sapiens, Rattus norvegicus, Mus musculus, Caenorhabditis elegans, Saccharomyces cerevisiae, Schizosaccharomyces pombe, and Candida albicans (Incyte Genomics, Palo Alto CA); hidden Markov model (HMM)-based protein family databases such as PFAM, INCY, and TIGRFAM (Haft, D.H. et al. (2001) Nucleic Acids Res. 29:41-43); and HMM-based protein domain databases such as SMART (Schultz et al. (1998) Proc. Natl. Acad. Sci. USA 95:5857-5864; Letunic, I. et al. (2002) Nucleic Acids Res. 30:242-244). (HMM is a probabilistic approach which analyzes consensus primary structures of gene families. See, for example, Eddy, S.R. (1996) Curr. Opin. Struct. Biol. 6:361-365.) The queries were performed using programs based on BLAST, FASTA, BLIMPS, and HMMER. The Incyte cDNA sequences were assembled to produce full length polynucleotide sequences. Alternatively, GenBank cDNAs, GenBank ESTs, stitched sequences, stretched sequences, or Genscan-predicted coding sequences (see Examples IV and V) were used to extend Incyte cDNA assemblages to full length. Assembly was performed using programs based on Phred, Phrap, and Consed, and cDNA assemblages were screened for open reading frames using programs based on GeneMark, BLAST, and FASTA. The full length polynucleotide sequences were translated to derive the corresponding full length polypeptide sequences. Alternatively, a polypeptide of the invention may begin at any of the methionine residues of the full length translated polypeptide. Full length polypeptide sequences were subsequently analyzed by querying against databases such as the GenBank protein databases (genpept), SwissProt, the PROTEOME databases, BLOCKS, PRINTS, DOMO, PRODOM, Prosite, hidden Markov model (HMM)-based protein family databases such as PFAM, INCY, and TIGRFAM; and HMM-based protein domain databases such as SMART. Full length polynucleotide sequences are also analyzed using MACDNASIS PRO software (Hitachi Software

Engineering, South San Francisco CA) and LASERGENE software (DNASTAR). Polynucleotide and polypeptide sequence alignments are generated using default parameters specified by the CLUSTAL algorithm as incorporated into the MEGALIGN multisequence alignment program (DNASTAR), which also calculates the percent identity between aligned sequences.

Table 7 summarizes the tools, programs, and algorithms used for the analysis and assembly of Incyte cDNA and full length sequences and provides applicable descriptions, references, and threshold parameters. The first column of Table 7 shows the tools, programs, and algorithms used, the second column provides brief descriptions thereof, the third column presents appropriate references, all of which are incorporated by reference herein in their entirety, and the fourth column presents, where applicable, the scores, probability values, and other parameters used to evaluate the strength of a match between two sequences (the higher the score or the lower the probability value, the greater the identity between two sequences).

The programs described above for the assembly and analysis of full length polynucleotide and polypeptide sequences were also used to identify polynucleotide sequence fragments from SEQ ID NO:13-24. Fragments from about 20 to about 4000 nucleotides which are useful in hybridization and amplification technologies are described in Table 4, column 2.

IV. Identification and Editing of Coding Sequences from Genomic DNA

Putative immunoglobulin superfamily proteins were initially identified by running the Genscan gene identification program against public genomic sequence databases (e.g., gbpri and gbhtg). Genscan is a general-purpose gene identification program which analyzes genomic DNA sequences from a variety of organisms (See Burge, C. and S. Karlin (1997) J. Mol. Biol. 268:78-94, and Burge, C. and S. Karlin (1998) Curr. Opin. Struct. Biol. 8:346-354). The program concatenates predicted exons to form an assembled cDNA sequence extending from a methionine to a stop codon. The output of Genscan is a FASTA database of polynucleotide and polypeptide sequences. The maximum range of sequence for Genscan to analyze at once was set to 30 kb. To determine which of these Genscan predicted cDNA sequences encode immunoglobulin superfamily proteins, the encoded polypeptides were analyzed by querying against PFAM models for immunoglobulin superfamily proteins. Potential immunoglobulin superfamily proteins were also identified by homology to Incyte cDNA sequences that had been annotated as immunoglobulin superfamily proteins. These selected Genscan-predicted sequences were then compared by BLAST analysis to the genpept and gbpri public databases. Where necessary, the Genscan-predicted sequences were then edited by comparison to the top BLAST hit from genpept to correct errors in the sequence predicted by Genscan, such as extra or omitted exons. BLAST analysis was also used to find any Incyte cDNA or

public cDNA coverage of the Genscan-predicted sequences, thus providing evidence for transcription. When Incyte cDNA coverage was available, this information was used to correct or confirm the Genscan predicted sequence. Full length polynucleotide sequences were obtained by assembling Genscan-predicted coding sequences with Incyte cDNA sequences and/or public cDNA sequences using the assembly process described in Example III. Alternatively, full length polynucleotide sequences were derived entirely from edited or unedited Genscan-predicted coding sequences.

V. Assembly of Genomic Sequence Data with cDNA Sequence Data

"Stitched" Sequences

Partial cDNA sequences were extended with exons predicted by the Genscan gene identification program described in Example IV. Partial cDNAs assembled as described in Example III were mapped to genomic DNA and parsed into clusters containing related cDNAs and Genscan exon predictions from one or more genomic sequences. Each cluster was analyzed using an algorithm based on graph theory and dynamic programming to integrate cDNA and genomic information, generating possible splice variants that were subsequently confirmed, edited, or extended to create a full length sequence. Sequence intervals in which the entire length of the interval was present on more than one sequence in the cluster were identified, and intervals thus identified were considered to be equivalent by transitivity. For example, if an interval was present on a cDNA and two genomic sequences, then all three intervals were considered to be equivalent. This process allows unrelated but consecutive genomic sequences to be brought together, bridged by cDNA sequence. Intervals thus identified were then "stitched" together by the stitching algorithm in the order that they appear along their parent sequences to generate the longest possible sequence, as well as sequence variants. Linkages between intervals which proceed along one type of parent sequence (cDNA to cDNA or genomic sequence to genomic sequence) were given preference over linkages which change parent type (cDNA to genomic sequence). The resultant stitched sequences were translated and compared by BLAST analysis to the genpept and gbpri public databases. Incorrect exons predicted by Genscan were corrected by comparison to the top BLAST hit from genpept. Sequences were further extended with additional cDNA sequences, or by inspection of genomic DNA, when necessary.

"Stretched" Sequences

Partial DNA sequences were extended to full length with an algorithm based on BLAST analysis. First, partial cDNAs assembled as described in Example III were queried against public databases such as the GenBank primate, rodent, mammalian, vertebrate, and eukaryote databases using the BLAST program. The nearest GenBank protein homolog was then compared by BLAST analysis to either Incyte cDNA sequences or GenScan exon predicted sequences described in

Example IV. A chimeric protein was generated by using the resultant high-scoring segment pairs (HSPs) to map the translated sequences onto the GenBank protein homolog. Insertions or deletions may occur in the chimeric protein with respect to the original GenBank protein homolog. The GenBank protein homolog, the chimeric protein, or both were used as probes to search for homologous genomic sequences from the public human genome databases. Partial DNA sequences were therefore "stretched" or extended by the addition of homologous genomic sequences. The resultant stretched sequences were examined to determine whether it contained a complete gene.

VI. Chromosomal Mapping of IGSFP Encoding Polynucleotides

The sequences which were used to assemble SEQ ID NO:13-24 were compared with sequences from the Incyte LIFESEQ database and public domain databases using BLAST and other implementations of the Smith-Waterman algorithm. Sequences from these databases that matched SEQ ID NO:13-24 were assembled into clusters of contiguous and overlapping sequences using assembly algorithms such as Phrap (Table 7). Radiation hybrid and genetic mapping data available from public resources such as the Stanford Human Genome Center (SHGC), Whitehead Institute for Genome Research (WIGR), and Généthon were used to determine if any of the clustered sequences had been previously mapped. Inclusion of a mapped sequence in a cluster resulted in the assignment of all sequences of that cluster, including its particular SEQ ID NO:, to that map location.

Map locations are represented by ranges, or intervals, of human chromosomes. The map position of an interval, in centiMorgans, is measured relative to the terminus of the chromosome's p-arm. (The centiMorgan (cM) is a unit of measurement based on recombination frequencies between chromosomal markers. On average, 1 cM is roughly equivalent to 1 megabase (Mb) of DNA in humans, although this can vary widely due to hot and cold spots of recombination.) The cM distances are based on genetic markers mapped by Généthon which provide boundaries for radiation hybrid markers whose sequences were included in each of the clusters. Human genome maps and other resources available to the public, such as the NCBI "GeneMap'99" World Wide Web site (http://www.ncbi.nlm.nih.gov/genemap/), can be employed to determine if previously identified disease genes map within or in proximity to the intervals indicated above.

VII. Analysis of Polynucleotide Expression

Northern analysis is a laboratory technique used to detect the presence of a transcript of a gene and involves the hybridization of a labeled nucleotide sequence to a membrane on which RNAs from a particular cell type or tissue have been bound. (See, e.g., Sambrook, supra, ch. 7; Ausubel (1995) supra, ch. 4 and 16.)

Analogous computer techniques applying BLAST were used to search for identical or related

molecules in cDNA databases such as GenBank or LIFESEQ (Incyte Genomics). This analysis is much faster than multiple membrane-based hybridizations. In addition, the sensitivity of the computer search can be modified to determine whether any particular match is categorized as exact or similar. The basis of the search is the product score, which is defined as:

$$\frac{\text{BLAST Score} \times \text{Percent Identity}}{5 \times \text{minimum } \{\text{length(Seq. 1)}, \text{length(Seq. 2)}\}}$$

The product score takes into account both the degree of similarity between two sequences and the length of the sequence match. The product score is a normalized value between 0 and 100, and is calculated as follows: the BLAST score is multiplied by the percent nucleotide identity and the product is divided by (5 times the length of the shorter of the two sequences). The BLAST score is calculated by assigning a score of +5 for every base that matches in a high-scoring segment pair (HSP), and -4 for every mismatch. Two sequences may share more than one HSP (separated by gaps). If there is more than one HSP, then the pair with the highest BLAST score is used to calculate the product score. The product score represents a balance between fractional overlap and quality in a BLAST alignment. For example, a product score of 100 is produced only for 100% identity over the entire length of the shorter of the two sequences being compared. A product score of 70 is produced either by 100% identity and 70% overlap at one end, or by 88% identity and 100% overlap at the other. A product score of 50 is produced either by 100% identity and 50% overlap at one end, or 79% identity and 100% overlap.

Alternatively, polynucleotide sequences encoding IGSFP are analyzed with respect to the tissue sources from which they were derived. For example, some full length sequences are assembled, at least in part, with overlapping Incyte cDNA sequences (see Example III). Each cDNA sequence is derived from a cDNA library constructed from a human tissue. Each human tissue is classified into one of the following organ/tissue categories: cardiovascular system; connective tissue; digestive system; embryonic structures; endocrine system; exocrine glands; genitalia, female; genitalia, male; germ cells; hemic and immune system; liver; musculoskeletal system; nervous system; pancreas; respiratory system; sense organs; skin; stomatognathic system; unclassified/mixed; or urinary tract. The number of libraries in each category is counted and divided by the total number of libraries across all categories. Similarly, each human tissue is classified into one of the following disease/condition categories: cancer, cell line, developmental, inflammation, neurological, trauma, cardiovascular, pooled, and other, and the number of libraries in each category is counted and divided

by the total number of libraries across all categories. The resulting percentages reflect the tissue- and disease-specific expression of cDNA encoding IGSFP. cDNA sequences and cDNA library/tissue information are found in the LIFESEQ GOLD database (Incyte Genomics, Palo Alto CA).

VIII. Extension of IGSFP Encoding Polynucleotides

Full length polynucleotide sequences were also produced by extension of an appropriate fragment of the full length molecule using oligonucleotide primers designed from this fragment. One primer was synthesized to initiate 5' extension of the known fragment, and the other primer was synthesized to initiate 3' extension of the known fragment. The initial primers were designed using OLIGO 4.06 software (National Biosciences), or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the target sequence at temperatures of about 68°C to about 72°C. Any stretch of nucleotides which would result in hairpin structures and primer-primer dimerizations was avoided.

Selected human cDNA libraries were used to extend the sequence. If more than one extension was necessary or desired, additional or nested sets of primers were designed.

High fidelity amplification was obtained by PCR using methods well known in the art. PCR was performed in 96-well plates using the PTC-200 thermal cycler (MJ Research, Inc.). The reaction mix contained DNA template, 200 nmol of each primer, reaction buffer containing Mg^{2+}, $(NH_4)_2SO_4$, and 2-mercaptoethanol, Taq DNA polymerase (Amersham Pharmacia Biotech), ELONGASE enzyme (Life Technologies), and Pfu DNA polymerase (Stratagene), with the following parameters for primer pair PCI A and PCI B: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68°C, 5 min; Step 7: storage at 4°C. In the alternative, the parameters for primer pair T7 and SK+ were as follows: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 57°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68°C, 5 min; Step 7: storage at 4°C.

The concentration of DNA in each well was determined by dispensing 100 µl PICOGREEN quantitation reagent (0.25% (v/v) PICOGREEN; Molecular Probes, Eugene OR) dissolved in 1X TE and 0.5 µl of undiluted PCR product into each well of an opaque fluorimeter plate (Corning Costar, Acton MA), allowing the DNA to bind to the reagent. The plate was scanned in a Fluoroskan II (Labsystems Oy, Helsinki, Finland) to measure the fluorescence of the sample and to quantify the concentration of DNA. A 5 µl to 10 µl aliquot of the reaction mixture was analyzed by electrophoresis on a 1 % agarose gel to determine which reactions were successful in extending the sequence.

The extended nucleotides were desalted and concentrated, transferred to 384-well plates,

digested with CviJI cholera virus endonuclease (Molecular Biology Research, Madison WI), and sonicated or sheared prior to religation into pUC 18 vector (Amersham Pharmacia Biotech). For shotgun sequencing, the digested nucleotides were separated on low concentration (0.6 to 0.8%) agarose gels, fragments were excised, and agar digested with Agar ACE (Promega). Extended clones were religated using T4 ligase (New England Biolabs, Beverly MA) into pUC 18 vector (Amersham Pharmacia Biotech), treated with Pfu DNA polymerase (Stratagene) to fill-in restriction site overhangs, and transfected into competent E. coli cells. Transformed cells were selected on antibiotic-containing media, and individual colonies were picked and cultured overnight at 37°C in 384-well plates in LB/2x carb liquid media.

The cells were lysed, and DNA was amplified by PCR using Taq DNA polymerase (Amersham Pharmacia Biotech) and Pfu DNA polymerase (Stratagene) with the following parameters: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 72°C, 2 min; Step 5: steps 2, 3, and 4 repeated 29 times; Step 6: 72°C, 5 min; Step 7: storage at 4°C. DNA was quantified by PICOGREEN reagent (Molecular Probes) as described above. Samples with low DNA recoveries were reamplified using the same conditions as described above. Samples were diluted with 20% dimethysulfoxide (1:2, v/v), and sequenced using DYENAMIC energy transfer sequencing primers and the DYENAMIC DIRECT kit (Amersham Pharmacia Biotech) or the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (Applied Biosystems).

In like manner, full length polynucleotide sequences are verified using the above procedure or are used to obtain 5' regulatory sequences using the above procedure along with oligonucleotides designed for such extension, and an appropriate genomic library.

IX. Identification of Single Nucleotide Polymorphisms in IGSFP Encoding Polynucleotides

Common DNA sequence variants known as single nucleotide polymorphisms (SNPs) were identified in SEQ ID NO:13-24 using the LIFESEQ database (Incyte Genomics). Sequences from the same gene were clustered together and assembled as described in Example III, allowing the identification of all sequence variants in the gene. An algorithm consisting of a series of filters was used to distinguish SNPs from other sequence variants. Preliminary filters removed the majority of basecall errors by requiring a minimum Phred quality score of 15, and removed sequence alignment errors and errors resulting from improper trimming of vector sequences, chimeras, and splice variants. An automated procedure of advanced chromosome analysis analysed the original chromatogram files in the vicinity of the putative SNP. Clone error filters used statistically generated algorithms to identify errors introduced during laboratory processing, such as those caused by reverse transcriptase,

polymerase, or somatic mutation. Clustering error filters used statistically generated algorithms to identify errors resulting from clustering of close homologs or pseudogenes, or due to contamination by non-human sequences. A final set of filters removed duplicates and SNPs found in immunoglobulins or T-cell receptors.

Certain SNPs were selected for further characterization by mass spectrometry using the high throughput MASSARRAY system (Sequenom, Inc.) to analyze allele frequencies at the SNP sites in four different human populations. The Caucasian population comprised 92 individuals (46 male, 46 female), including 83 from Utah, four French, three Venezualan, and two Amish individuals. The African population comprised 194 individuals (97 male, 97 female), all African Americans. The Hispanic population comprised 324 individuals (162 male, 162 female), all Mexican Hispanic. The Asian population comprised 126 individuals (64 male, 62 female) with a reported parental breakdown of 43% Chinese, 31% Japanese, 13% Korean, 5% Vietnamese, and 8% other Asian. Allele frequencies were first analyzed in the Caucasian population; in some cases those SNPs which showed no allelic variance in this population were not further tested in the other three populations.

X. Labeling and Use of Individual Hybridization Probes

Hybridization probes derived from SEQ ID NO:13-24 are employed to screen cDNAs, genomic DNAs, or mRNAs. Although the labeling of oligonucleotides, consisting of about 20 base pairs, is specifically described, essentially the same procedure is used with larger nucleotide fragments. Oligonucleotides are designed using state-of-the-art software such as OLIGO 4.06 software (National Biosciences) and labeled by combining 50 pmol of each oligomer, 250 μCi of [γ-^{32}P] adenosine triphosphate (Amersham Pharmacia Biotech), and T4 polynucleotide kinase (DuPont NEN, Boston MA). The labeled oligonucleotides are substantially purified using a SEPHADEX G-25 superfine size exclusion dextran bead column (Amersham Pharmacia Biotech). An aliquot containing 10^7 counts per minute of the labeled probe is used in a typical membrane-based hybridization analysis of human genomic DNA digested with one of the following endonucleases: Ase I, Bgl II, Eco RI, Pst I, Xba I, or Pvu II (DuPont NEN).

The DNA from each digest is fractionated on a 0.7% agarose gel and transferred to nylon membranes (Nytran Plus, Schleicher & Schuell, Durham NH). Hybridization is carried out for 16 hours at 40°C. To remove nonspecific signals, blots are sequentially washed at room temperature under conditions of up to, for example, 0.1 x saline sodium citrate and 0.5% sodium dodecyl sulfate. Hybridization patterns are visualized using autoradiography or an alternative imaging means and compared.

XI. Microarrays

The linkage or synthesis of array elements upon a microarray can be achieved utilizing photolithography, piezoelectric printing (ink-jet printing, See, e.g., Baldeschweiler, supra.), mechanical microspotting technologies, and derivatives thereof. The substrate in each of the aforementioned technologies should be uniform and solid with a non-porous surface (Schena (1999), supra). Suggested substrates include silicon, silica, glass slides, glass chips, and silicon wafers. Alternatively, a procedure analogous to a dot or slot blot may also be used to arrange and link elements to the surface of a substrate using thermal, UV, chemical, or mechanical bonding procedures. A typical array may be produced using available methods and machines well known to those of ordinary skill in the art and may contain any appropriate number of elements. (See, e.g., Schena, M. et al. (1995) Science 270:467-470; Shalon, D. et al. (1996) Genome Res. 6:639-645; Marshall, A. and J. Hodgson (1998) Nat. Biotechnol. 16:27-31.)

Full length cDNAs, Expressed Sequence Tags (ESTs), or fragments or oligomers thereof may comprise the elements of the microarray. Fragments or oligomers suitable for hybridization can be selected using software well known in the art such as LASERGENE software (DNASTAR). The array elements are hybridized with polynucleotides in a biological sample. The polynucleotides in the biological sample are conjugated to a fluorescent label or other molecular tag for ease of detection. After hybridization, nonhybridized nucleotides from the biological sample are removed, and a fluorescence scanner is used to detect hybridization at each array element. Alternatively, laser desorbtion and mass spectrometry may be used for detection of hybridization. The degree of complementarity and the relative abundance of each polynucleotide which hybridizes to an element on the microarray may be assessed. In one embodiment, microarray preparation and usage is described in detail below.

Tissue or Cell Sample Preparation

Total RNA is isolated from tissue samples using the guanidinium thiocyanate method and poly(A)$^+$ RNA is purified using the oligo-(dT) cellulose method. Each poly(A)$^+$ RNA sample is reverse transcribed using MMLV reverse-transcriptase, 0.05 pg/μl oligo-(dT) primer (21mer), 1X first strand buffer, 0.03 units/μl RNase inhibitor, 500 μM dATP, 500 μM dGTP, 500 μM dTTP, 40 μM dCTP, 40 μM dCTP-Cy3 (BDS) or dCTP-Cy5 (Amersham Pharmacia Biotech). The reverse transcription reaction is performed in a 25 ml volume containing 200 ng poly(A)$^+$ RNA with GEMBRIGHT kits (Incyte). Specific control poly(A)$^+$ RNAs are synthesized by in vitro transcription from non-coding yeast genomic DNA. After incubation at 37°C for 2 hr, each reaction sample (one with Cy3 and another with Cy5 labeling) is treated with 2.5 ml of 0.5M sodium hydroxide and incubated for 20 minutes at 85°C to the stop the reaction and degrade the RNA. Samples are purified

using two successive CHROMA SPIN 30 gel filtration spin columns (CLONTECH Laboratories, Inc. (CLONTECH), Palo Alto CA) and after combining, both reaction samples are ethanol precipitated using 1 ml of glycogen (1 mg/ml), 60 ml sodium acetate, and 300 ml of 100% ethanol. The sample is then dried to completion using a SpeedVAC (Savant Instruments Inc., Holbrook NY) and resuspended in 14 μl 5X SSC/0.2% SDS.

Microarray Preparation

Sequences of the present invention are used to generate array elements. Each array element is amplified from bacterial cells containing vectors with cloned cDNA inserts. PCR amplification uses primers complementary to the vector sequences flanking the cDNA insert. Array elements are amplified in thirty cycles of PCR from an initial quantity of 1-2 ng to a final quantity greater than 5 μg. Amplified array elements are then purified using SEPHACRYL-400 (Amersham Pharmacia Biotech).

Purified array elements are immobilized on polymer-coated glass slides. Glass microscope slides (Corning) are cleaned by ultrasound in 0.1% SDS and acetone, with extensive distilled water washes between and after treatments. Glass slides are etched in 4% hydrofluoric acid (VWR Scientific Products Corporation (VWR), West Chester PA), washed extensively in distilled water, and coated with 0.05% aminopropyl silane (Sigma) in 95% ethanol. Coated slides are cured in a 110°C oven.

Array elements are applied to the coated glass substrate using a procedure described in U.S. Patent No. 5,807,522, incorporated herein by reference. 1 μl of the array element DNA, at an average concentration of 100 ng/μl, is loaded into the open capillary printing element by a high-speed robotic apparatus. The apparatus then deposits about 5 nl of array element sample per slide.

Microarrays are UV-crosslinked using a STRATALINKER UV-crosslinker (Stratagene). Microarrays are washed at room temperature once in 0.2% SDS and three times in distilled water. Non-specific binding sites are blocked by incubation of microarrays in 0.2% casein in phosphate buffered saline (PBS) (Tropix, Inc., Bedford MA) for 30 minutes at 60°C followed by washes in 0.2% SDS and distilled water as before.

Hybridization

Hybridization reactions contain 9 μl of sample mixture consisting of 0.2 μg each of Cy3 and Cy5 labeled cDNA synthesis products in 5X SSC, 0.2% SDS hybridization buffer. The sample mixture is heated to 65°C for 5 minutes and is aliquoted onto the microarray surface and covered with an 1.8 cm^2 coverslip. The arrays are transferred to a waterproof chamber having a cavity just slightly larger than a microscope slide. The chamber is kept at 100% humidity internally by the addition of 140 μl of 5X SSC in a corner of the chamber. The chamber containing the arrays is incubated for about

6.5 hours at 60° C. The arrays are washed for 10 min at 45°C in a first wash buffer (1X SSC, 0.1% SDS), three times for 10 minutes each at 45°C in a second wash buffer (0.1X SSC), and dried.

Detection

Reporter-labeled hybridization complexes are detected with a microscope equipped with an Innova 70 mixed gas 10 W laser (Coherent, Inc., Santa Clara CA) capable of generating spectral lines at 488 nm for excitation of Cy3 and at 632 nm for excitation of Cy5. The excitation laser light is focused on the array using a 20X microscope objective (Nikon, Inc., Melville NY). The slide containing the array is placed on a computer-controlled X-Y stage on the microscope and raster-scanned past the objective. The 1.8 cm x 1.8 cm array used in the present example is scanned with a resolution of 20 micrometers.

In two separate scans, a mixed gas multiline laser excites the two fluorophores sequentially. Emitted light is split, based on wavelength, into two photomultiplier tube detectors (PMT R1477, Hamamatsu Photonics Systems, Bridgewater NJ) corresponding to the two fluorophores. Appropriate filters positioned between the array and the photomultiplier tubes are used to filter the signals. The emission maxima of the fluorophores used are 565 nm for Cy3 and 650 nm for Cy5. Each array is typically scanned twice, one scan per fluorophore using the appropriate filters at the laser source, although the apparatus is capable of recording the spectra from both fluorophores simultaneously.

The sensitivity of the scans is typically calibrated using the signal intensity generated by a cDNA control species added to the sample mixture at a known concentration. A specific location on the array contains a complementary DNA sequence, allowing the intensity of the signal at that location to be correlated with a weight ratio of hybridizing species of 1:100,000. When two samples from different sources (e.g., representing test and control cells), each labeled with a different fluorophore, are hybridized to a single array for the purpose of identifying genes that are differentially expressed, the calibration is done by labeling samples of the calibrating cDNA with the two fluorophores and adding identical amounts of each to the hybridization mixture.

The output of the photomultiplier tube is digitized using a 12-bit RTI-835H analog-to-digital (A/D) conversion board (Analog Devices, Inc., Norwood MA) installed in an IBM-compatible PC computer. The digitized data are displayed as an image where the signal intensity is mapped using a linear 20-color transformation to a pseudocolor scale ranging from blue (low signal) to red (high signal). The data is also analyzed quantitatively. Where two different fluorophores are excited and measured simultaneously, the data are first corrected for optical crosstalk (due to overlapping emission spectra) between the fluorophores using each fluorophore's emission spectrum.

A grid is superimposed over the fluorescence signal image such that the signal from each spot

is centered in each element of the grid. The fluorescence signal within each element is then integrated to obtain a numerical value corresponding to the average intensity of the signal. The software used for signal analysis is the GEMTOOLS gene expression analysis program (Incyte).

For example, SEQ ID NO:19 showed differential expression in toxicology studies as determined by microarray analysis. The expression of SEQ ID NO:19 was decreased by at least two fold in a human C3A liver cell line treated with various drugs (e.g., steroids, steroid hormones) relative to untreated C3A cells. The human C3A cell line is a clonal derivative of HepG2/C3 (hepatoma cell line, isolated from a 15-year-old male with liver tumor), which was selected for strong contact inhibition of growth. The C3A cell line is well established as an in vitro model of the mature human liver (Mickelson et al. (1995) Hepatology 22:866-875; Nagendra et al. (1997) Am J Physiol 272:G408-G416). Effects upon liver metabolism are important to understanding the pharmacodynamics of a drug. Therefore, SEQ ID NO:19 is useful for understanding the pharmacodynamics of a drug.

XII. Complementary Polynucleotides

Sequences complementary to the IGSFP-encoding sequences, or any parts thereof, are used to detect, decrease, or inhibit expression of naturally occurring IGSFP. Although use of oligonucleotides comprising from about 15 to 30 base pairs is described, essentially the same procedure is used with smaller or with larger sequence fragments. Appropriate oligonucleotides are designed using OLIGO 4.06 software (National Biosciences) and the coding sequence of IGSFP. To inhibit transcription, a complementary oligonucleotide is designed from the most unique 5' sequence and used to prevent promoter binding to the coding sequence. To inhibit translation, a complementary oligonucleotide is designed to prevent ribosomal binding to the IGSFP-encoding transcript.

XIII. Expression of IGSFP

Expression and purification of IGSFP is achieved using bacterial or virus-based expression systems. For expression of IGSFP in bacteria, cDNA is subcloned into an appropriate vector containing an antibiotic resistance gene and an inducible promoter that directs high levels of cDNA transcription. Examples of such promoters include, but are not limited to, the *trp-lac* (*tac*) hybrid promoter and the T5 or T7 bacteriophage promoter in conjunction with the *lac* operator regulatory element. Recombinant vectors are transformed into suitable bacterial hosts, e.g., BL21(DE3). Antibiotic resistant bacteria express IGSFP upon induction with isopropyl beta-D-thiogalactopyranoside (IPTG). Expression of IGSFP in eukaryotic cells is achieved by infecting insect or mammalian cell lines with recombinant Autographica californica nuclear polyhedrosis virus (AcMNPV), commonly known as baculovirus. The nonessential polyhedrin gene of baculovirus is replaced with cDNA encoding IGSFP by either homologous recombination or bacterial-mediated

transposition involving transfer plasmid intermediates. Viral infectivity is maintained and the strong polyhedrin promoter drives high levels of cDNA transcription. Recombinant baculovirus is used to infect Spodoptera frugiperda (Sf9) insect cells in most cases, or human hepatocytes, in some cases. Infection of the latter requires additional genetic modifications to baculovirus. (See Engelhard, E.K. et al. (1994) Proc. Natl. Acad. Sci. USA 91:3224-3227; Sandig, V. et al. (1996) Hum. Gene Ther. 7:1937-1945.)

In most expression systems, IGSFP is synthesized as a fusion protein with, e.g., glutathione S-transferase (GST) or a peptide epitope tag, such as FLAG or 6-His, permitting rapid, single-step, affinity-based purification of recombinant fusion protein from crude cell lysates. GST, a 26-kilodalton enzyme from Schistosoma japonicum, enables the purification of fusion proteins on immobilized glutathione under conditions that maintain protein activity and antigenicity (Amersham Pharmacia Biotech). Following purification, the GST moiety can be proteolytically cleaved from IGSFP at specifically engineered sites. FLAG, an 8-amino acid peptide, enables immunoaffinity purification using commercially available monoclonal and polyclonal anti-FLAG antibodies (Eastman Kodak). 6-His, a stretch of six consecutive histidine residues, enables purification on metal-chelate resins (QIAGEN). Methods for protein expression and purification are discussed in Ausubel (1995, supra, ch. 10 and 16). Purified IGSFP obtained by these methods can be used directly in the assays shown in Examples XVII and XVIII where applicable.

XIV. Functional Assays

IGSFP function is assessed by expressing the sequences encoding IGSFP at physiologically elevated levels in mammalian cell culture systems. cDNA is subcloned into a mammalian expression vector containing a strong promoter that drives high levels of cDNA expression. Vectors of choice include PCMV SPORT (Life Technologies) and PCR3.1 (Invitrogen, Carlsbad CA), both of which contain the cytomegalovirus promoter. 5-10 μg of recombinant vector are transiently transfected into a human cell line, for example, an endothelial or hematopoietic cell line, using either liposome formulations or electroporation. 1-2 μg of an additional plasmid containing sequences encoding a marker protein are co-transfected. Expression of a marker protein provides a means to distinguish transfected cells from nontransfected cells and is a reliable predictor of cDNA expression from the recombinant vector. Marker proteins of choice include, e.g., Green Fluorescent Protein (GFP; Clontech), CD64, or a CD64-GFP fusion protein. Flow cytometry (FCM), an automated, laser optics-based technique, is used to identify transfected cells expressing GFP or CD64-GFP and to evaluate the apoptotic state of the cells and other cellular properties. FCM detects and quantifies the uptake of fluorescent molecules that diagnose events preceding or coincident with cell death. These events

include changes in nuclear DNA content as measured by staining of DNA with propidium iodide; changes in cell size and granularity as measured by forward light scatter and 90 degree side light scatter; down-regulation of DNA synthesis as measured by decrease in bromodeoxyuridine uptake; alterations in expression of cell surface and intracellular proteins as measured by reactivity with specific antibodies; and alterations in plasma membrane composition as measured by the binding of fluorescein-conjugated Annexin V protein to the cell surface. Methods in flow cytometry are discussed in Ormerod, M.G. (1994) Flow Cytometry, Oxford, New York NY.

The influence of IGSFP on gene expression can be assessed using highly purified populations of cells transfected with sequences encoding IGSFP and either CD64 or CD64-GFP. CD64 and CD64-GFP are expressed on the surface of transfected cells and bind to conserved regions of human immunoglobulin G (IgG). Transfected cells are efficiently separated from nontransfected cells using magnetic beads coated with either human IgG or antibody against CD64 (DYNAL, Lake Success NY). mRNA can be purified from the cells using methods well known by those of skill in the art. Expression of mRNA encoding IGSFP and other genes of interest can be analyzed by northern analysis or microarray techniques.

XV. Production of IGSFP Specific Antibodies

IGSFP substantially purified using polyacrylamide gel electrophoresis (PAGE; see, e.g., Harrington, M.G. (1990) Methods Enzymol. 182:488-495), or other purification techniques, is used to immunize animals (e.g., rabbits, mice, etc.) and to produce antibodies using standard protocols.

Alternatively, the IGSFP amino acid sequence is analyzed using LASERGENE software (DNASTAR) to determine regions of high immunogenicity, and a corresponding oligopeptide is synthesized and used to raise antibodies by means known to those of skill in the art. Methods for selection of appropriate epitopes, such as those near the C-terminus or in hydrophilic regions are well described in the art. (See, e.g., Ausubel, 1995, supra, ch. 11.)

Typically, oligopeptides of about 15 residues in length are synthesized using an ABI 431A peptide synthesizer (Applied Biosystems) using FMOC chemistry and coupled to KLH (Sigma-Aldrich, St. Louis MO) by reaction with N-maleimidobenzoyl-N-hydroxysuccinimide ester (MBS) to increase immunogenicity. (See, e.g., Ausubel, 1995, supra.) Rabbits are immunized with the oligopeptide-KLH complex in complete Freund's adjuvant. Resulting antisera are tested for antipeptide and anti-IGSFP activity by, for example, binding the peptide or IGSFP to a substrate, blocking with 1% BSA, reacting with rabbit antisera, washing, and reacting with radio-iodinated goat anti-rabbit IgG.

XVI. Purification of Naturally Occurring IGSFP Using Specific Antibodies

Naturally occurring or recombinant IGSFP is substantially purified by immunoaffinity chromatography using antibodies specific for IGSFP. An immunoaffinity column is constructed by covalently coupling anti-IGSFP antibody to an activated chromatographic resin, such as CNBr-activated SEPHAROSE (Amersham Pharmacia Biotech). After the coupling, the resin is blocked and washed according to the manufacturer's instructions.

Media containing IGSFP are passed over the immunoaffinity column, and the column is washed under conditions that allow the preferential absorbance of IGSFP (e.g., high ionic strength buffers in the presence of detergent). The column is eluted under conditions that disrupt antibody/IGSFP binding (e.g., a buffer of pH 2 to pH 3, or a high concentration of a chaotrope, such as urea or thiocyanate ion), and IGSFP is collected.

XVII. Identification of Molecules Which Interact with IGSFP

IGSFP, or biologically active fragments thereof, are labeled with ^{125}I Bolton-Hunter reagent. (See, e.g., Bolton, A.E. and W.M. Hunter (1973) Biochem. J. 133:529-539.) Candidate molecules previously arrayed in the wells of a multi-well plate are incubated with the labeled IGSFP, washed, and any wells with labeled IGSFP complex are assayed. Data obtained using different concentrations of IGSFP are used to calculate values for the number, affinity, and association of IGSFP with the candidate molecules.

Alternatively, molecules interacting with IGSFP are analyzed using the yeast two-hybrid system as described in Fields, S. and O. Song (1989) Nature 340:245-246, or using commercially available kits based on the two-hybrid system, such as the MATCHMAKER system (Clontech).

IGSFP may also be used in the PATHCALLING process (CuraGen Corp., New Haven CT) which employs the yeast two-hybrid system in a high-throughput manner to determine all interactions between the proteins encoded by two large libraries of genes (Nandabalan, K. et al. (2000) U.S. Patent No. 6,057,101).

XVIII. Demonstration of IGSFP Activity

An assay for IGSFP activity measures the ability of IGSFP to recognize and precipitate antigens from serum. This activity can be measured by the quantitative precipitin reaction. (Golub, E. S. et al. (1987) Immunology: A Synthesis, Sinauer Associates, Sunderland, MA, pages 113-115.) IGSFP is isotopically labeled using methods known in the art. Various serum concentrations are added to constant amounts of labeled IGSFP. IGSFP-antigen complexes precipitate out of solution and are collected by centrifugation. The amount of precipitable IGSFP-antigen complex is proportional to the amount of radioisotope detected in the precipitate. The amount of precipitable

IGSFP-antigen complex is plotted against the serum concentration. For various serum concentrations, a characteristic precipitin curve is obtained, in which the amount of precipitable IGSFP-antigen complex initially increases proportionately with increasing serum concentration, peaks at the equivalence point, and then decreases proportionately with further increases in serum concentration. Thus, the amount of precipitable IGSFP-antigen complex is a measure of IGSFP activity which is characterized by sensitivity to both limiting and excess quantities of antigen.

Alternatively, an assay for IGSFP activity measures the expression of IGSFP on the cell surface. cDNA encoding IGSFP is transfected into a non-leukocytic cell line. Cell surface proteins are labeled with biotin (de la Fuente, M.A. et.al. (1997) Blood 90:2398-2405). Immunoprecipitations are performed using IGSFP-specific antibodies, and immunoprecipitated samples are analyzed using SDS-PAGE and immunoblotting techniques. The ratio of labeled immunoprecipitant to unlabeled immunoprecipitant is proportional to the amount of IGSFP expressed on the cell surface.

Alternatively, an assay for IGSFP activity measures the amount of cell aggregation induced by overexpression of IGSFP. In this assay, cultured cells such as NIH3T3 are transfected with cDNA encoding IGSFP contained within a suitable mammalian expression vector under control of a strong promoter. Cotransfection with cDNA encoding a fluorescent marker protein, such as Green Fluorescent Protein (CLONTECH), is useful for identifying stable transfectants. The amount of cell agglutination, or clumping, associated with transfected cells is compared with that associated with untransfected cells. The amount of cell agglutination is a direct measure of IGSFP activity.

Various modifications and variations of the described methods and systems of the invention will be apparent to those skilled in the art without departing from the scope and spirit of the invention. Although the invention has been described in connection with certain embodiments, it should be understood that the invention as claimed should not be unduly limited to such specific embodiments. Indeed, various modifications of the described modes for carrying out the invention which are obvious to those skilled in molecular biology or related fields are intended to be within the scope of the following claims.

Table 1

Incyte Project ID	Polypeptide SEQ ID NO:	Incyte Polypeptide ID	Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID	CA2 Reagents
3855123	1	3855123CD1	13	3855123CB1	
4547188	2	4547188CD1	14	4547188CB1	90065916CA2
3939883	3	3939883CD1	15	3939883CB1	
3163819	4	3163819CD1	16	3163819CB1	3163819CA2
8518269	5	8518269CD1	17	8518269CB1	90110559CA2
1592646	6	1592646CD1	18	1592646CB1	
7500191	7	7500191CD1	19	7500191CB1	
7500099	8	7500099CD1	20	7500099CB1	2836421CA2
7682434	9	7682434CD1	21	7682434CB1	
2202389	10	2202389CD1	22	2202389CB1	
7503597	11	7503597CD1	23	7503597CB1	
7503603	12	7503603CD1	24	7503603CB1	

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
1	3855123CD1	g14572521	2.00E-70	[Homo sapiens] NEPH1 (Donoviel,D.B. et al. (2001) Mol. Cell. Biol. 21 (14), 4829-4836)
2	4547188CD1	g11071950	9.60E-121	[Mus musculus] (AB048834) Fca/m receptor (Shibuya,A. et al. (2001) Nat. Immunol. 1 (5), 441-446)
3	3939883CD1	g1136501	6.90E-35	[Rattus norvegicus] surface protein MCA-32 (Pirozzi,G. et al. (1995) J. Immunol. 155 (12), 5811-5818)
4	3163819CD1	g9887089	6.50E-32	[Mus musculus] lymphocyte antigen 108 isoform 1 (Peck,S.R. et al. (2000) Immunogenetics 52 (1-2), 63-72)
4	3163819CD1	g15384841	1.00E-112	[Homo sapiens] activating NK receptor (Bottino,C. et al. (2001) The Journal of experimental medicine. 194 (3), 235-246)
5	8518269CD1	g9887089	5.20E-62	[Mus musculus] lymphocyte antigen 108 isoform 1 (Peck,S.R. et al. (2000) Immunogenetics 52 (1-2), 63-72)
5	8518269CD1	g15384841	0.00E+00	[Homo sapiens] activating NK receptor (Bottino,C. et al. (2001) The Journal of experimental medicine. 194 (3), 235-246)
6	1592646CD1	g18376829	1.00E-154	[Homo sapiens] (AF391163) osteoclast-associated receptor hOSCAR-M2 (Kim,N. et al. (2002) J. Exp. Med. 195 (2), 201-209)
6	1592646CD1	g2645890	2.00E-32	[Homo sapiens] IGSF1 (Mazzarella,R. et al. (1998) Genomics 48 (2), 157-162)
7	7500191CD1	g2078518	0	[Homo sapiens] neogenin (Vielmetter,J. et al. (1997) Genomics 41 (3), 414-421)
8	7500099CD1	g10197717	7.40E-191	[Homo sapiens] cell-surface molecule Ly-9 (Tovar,V. et al. (2000) Immunogenetics 51 (10), 788-793)
9	7682434CD1	g586	1.20E-80	[Bos taurus] put. pre-OPCAM (AA 1 - 345) (Schofield,P.R. et al. (1989) EMBO J. 8 (2), 489-495)

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
		336698|OPCML	2.2E-81	[Homo sapiens][Receptor (signaling)][Plasma membrane] Opioid-binding cell adhesion molecule, has strong similarity to ratRn.11366, which is a glycosylphosphatidylinositol (GPI)-anchored neural cell adhesion molecule and a member of the immunoglobulin superfamily (Struyk, A. F., et al. (1995) Cloning of neurotrimin defines a new subfamily of differentially expressed neural cell adhesion molecules. J. Neurosci 15:2141-2156; Lane, C. M. et al. (1992) Regulation of an opioid-binding protein in NG108-15 cells parallels regulation of delta-opioid receptors. Proc Natl Acad Sci U S A 89:11234-11238.)
		332056|Rn.11366	5.3E-80	[Rattus norvegicus][Receptor (signaling)][Plasma membrane] Opioid-binding cell adhesion molecule, member of the immunoglobulin superfamily and a glycosylphosphatidylinositol (GPI)-anchored neural cell adhesion molecule (Hachisuka, A., et al. (2000) Developmental expression of opioid-binding cell adhesion molecule (OBCAM) in rat brain Brain Res Dev Brain Res 122:183-191.)
		330088|Lsamp	3.9E-77	[Rattus norvegicus][Plasma membrane] Limbic system-associated membrane protein, a member of the Ig family of proteins that plays a role in the selective growth of neurons and the targeting of axons (Pimenta, A. F., et al. (1996) cDNA cloning and structural analysis of the human limbic system-associated membrane protein (LAMP). Gene 170:189-195.)
11	7503597CD1	g14572521	5.40E-158	[Homo sapiens] NEPH1 (Donoviel,D.B. et al. (2001) Mol. Cell. Biol. 21 (14), 4829-4836)

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
		598720|FLJ10845	6.6E-66	[Homo sapiens] Protein containing an immunoglobulin (Ig) domain, has a region of low similarity to a region of raf Rn.11366, opioid-binding cell adhesion molecule, which is a glycosylphosphatidylinositol (GPI)-anchored neural cell adhesion molecule
		340970|NPHS1	5.6E-21	[Homo sapiens][Plasma membrane;Cell junction] Nephrin, a member of the immunoglobulin family expressed in renal glomeruli, may have a role in the development or function of the kidney filtration barrier; mutation of corresponding gene causes congenital nephrotic syndrome (Ruotsalainen, V. et al. (2000) Role of nephrin in cell junction formation in human nephrogenesis. Am. J. Pathol. 157:1905-1916.)

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
1	3855123CD1	442	S37 S51 S118 S129 S138 S171 S227 S236 S252 S366 S379 S385 S398 T48 T261 T306 T389 Y60	N162	Signal Peptide: M198-C223	HMMER
					Immunoglobulin domain: G13-A64, G97-A165	HMMER_PFAM
					Transmembrane domain: A193-A221 N-terminus is non-cytosolic	TMAP
					IMMUNOGLOBULIN DM00001|Q08180|426-518: V80-D172	BLAST_DOMO
2	4547188CD1	577	S39 S108 S189 S296 S301 S405 S482 S493 S525 T6 T38 T88 T234 T260 T271 T335 T349 T350 T437 T486 T524 T569 Y24	N212 N321	Signal Peptide: M46-P63, M46-Q64, P33-P63	HMMER
					Immunoglobulin domain: G120-I200	HMMER_PFAM
					Transmembrane domain: S39-P67 R495-R517 N-terminus is cytosolic	TMAP
					IMMUNOGLOBULIN DM00001|P01833|41-120: H128-G201	BLAST_DOMO
					IMMUNOGLOBULIN DM00001|P15083|41-120: H128-F208	BLAST_DOMO
					IMMUNOGLOBULIN DM00001|P01832|28-125: G120-G201	BLAST_DOMO

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
					IMMUNOGLOBULIN DM00001|S48841|41-120: H128-G201	BLAST_DOMO
3	3939883CD1	385	S4 S21 S99 S133 S214 S330 S373 T40 T60 T116 T162 T179 T181 T201 T226 T259 T296 T311 T342 Y236 Y355	N93 N102 N131 N193 N199 N224	signal_cleavage: M1-T38	SPSCAN
					Signal Peptide: M1-G41	HMMER
					Intracellular domains: M1-K19, K293-F385	TMHMMER
					Transmembrane domains: F20-S39, L270-P292	TMHMMER
					Extracellular domain: T40-K269	TMHMMER
					Immunoglobulin domain: G91-A147, D182-A240	HMMER_PFAM
					Receptor Fc Immunoglobulin PD01270: T135-V171, R183-P211 P value < 1.3e-3	BLIMPS_PRODOM
					SURFACE PROTEIN MCA32 PD095298: L30-V164	BLAST_PRODOM
					PLATELET ENDOTHELIAL CELL ADHESION PRECURSOR SIGNAL MOLECULE PECAM1 CD31 ANTIGEN PD150932: C68-P305	BLAST_PRODOM
					Leucine zipper pattern: L270-L291	MOTIFS
					Cell attachment sequence: R308-D310	MOTIFS
4	3163819CD1	221	S43 S52 S78 S143 S157 S180 T148 T188 T215 Y82	N26 N33 N50 N67 N92 N170 N192 N202	Signal Peptides: M1-G15, M1-L19, M1-N21	HMMER
					Extracellular domain: M1-K114	TMHMMER
					Transmembrane domain: M115-L137	TMHMMER
					Intracellular domain: R138-V221	TMHMMER

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
5	8518269CD1	332	S106 S112 S116 S154 S163 S189 S254 S268 S291 T123 T259 T299 T326 Y107 Y193	N58 N87 N137 N144 N161 N178 N203 N281 N303 N313	signal_cleavage: M1-S21	SPSCAN
					Signal Peptides: M1-G15, M1-V19, M1-S21, M1-S23	HMMER
					Extracellular domain: M1-K225	TMHMMER
					Transmembrane domain: M226-L248	TMHMMER
					Intracellular domain: R249-V332	TMHMMER
					Immunoglobulin domain: G35-I111, T146-A197	HMMER_PFAM
					ANTIGEN PRECURSOR SIGNAL IMMUNOGLOBULIN FOLD GLYCOPROTEIN TCELL SURFACE CD2 TRANSMEMBRANE PD010953: G32-S205	BLAST_PRODOM
6	1592646CD1	288	S122 S172 S232 S241 T75	N73 N170 N181	signal_cleavage: M1-T43	SPSCAN
					Signal Peptides: M26-T43, M1-T43	HMMER
					Immunoglobulin domain: G168-Y222, G71-Y127	HMMER_PFAM
					RECEPTOR NK CELL KILLER PRECURSOR SIGNAL LEUCOCYTE IMMUNOGLOBULIN-LIKE NATURAL INHIBITORY PD000659: H55-A193	BLAST_PRODOM
					ALPHA1BGLYCOPROTEIN IMMUNOGLOBULIN FOLD GLYCOPROTEIN PLASMA PD138678: Y54-I240	BLAST_PRODOM

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
7	7500191CD1	1450	S46 S64 S81 S156 S294 S451 S606 S620 S677 S731 S834 S939 S1087 S1137 S1203 S1281 S1283 S1291 S1327 S1328 S1385 S1407 S1423 T143 T212 T279 T311 T365 T371 T458 T532 T581 T603 T628 T759 T784 T808 T869 T873 T892 T924 T948 T1051	N73 N210 N326 N470 N489 N639 N715 N909 N1135 N1287	signal_cleavage: M1-A33	SPSCAN
			T1117 T1121 T1187 T1414 Y127 Y408 Y890		Signal Peptides: M1-G30, M1-A33	HMMER
					Fibronectin type III domain: P539-T621, P633-T721, P954-S1044, P439-S525, P739-L821, P853-S942	HMMER_PFAM
					Immunoglobulin domain: G263-A322, G166-V223, G67-A131, S355-A412	HMMER_PFAM
					Cytosolic domain: T1117-A1450 Transmembrane domain: L1094-C1116 Non-cytosolic domain: M1-M1093	TMHMMER
					Receptor tyrosine kinase class V proteins BL00790: V450-F476, Y477-G520, S554-K579	BLIMPS_BLOCKS

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
					Fibronectin type III repeat signature PR00014: T752-P761, A908-Y926, Y1028-P1042	BLIMPS_PRINTS
					TUMOR SUPPRESSOR NEOGENIN PROTEIN DCC PRECURSOR COLORECTAL GLYCOPROTEIN IMMUNOGLOBULIN FOLD PD041287: D1169-T1448 PD009999: C1116-P1172	BLAST_PRODOM
					NEOGENIN PROTEIN PD020198: M1-R66	BLAST_PRODOM
					TUMOR SUPPRESSOR PD171136: E58-V133	BLAST_PRODOM
					IMMUNOGLOBULIN DM00001|P43146|328-410: P341-Q420 DM00001|P43146|42-127: F55-I140	BLAST_DOMO
					FIBRONECTIN TYPE III REPEAT DM00007|P43146|935-1014: A945-D1025 DM00007|P43146|834-912: T846-N923	BLAST_DOMO
					TonB-dependent receptor proteins signature 1: M1-R5	MOTIFS
8	7500099CD1	551	S6 S17 S46 S128 S163 S179 S229 S316 S321 S400 S431 S453 S512 S524 T73 T122 T141 T142 T160 T192 T212 T252 T277 T438 T439 T487 Y335	N68 N95 N120 N169 N173 N285 N436	signal_cleavage: M1-G47	SPSCAN
					Immunoglobulin domain: S171-A224, G60-I133	HMMER_PFAM

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
					Cytosolic domain: K387-T551 Transmembrane domain: L365-W386 Non-cytosolic domain: M1-K364	TMHMMER
					ANTIGEN LY9 PRECURSOR SIGNAL TRANSMEMBRANE GLYCOPROTEIN IMMUNOGLOBULIN FOLD PD126134: P359-P545	BLAST_PRODOM
					ANTIGEN PRECURSOR SIGNAL IMMUNOGLOBULIN FOLD GLYCOPROTEIN TCELL SURFACE CD2 TRANSMEMBRANE PD010953: V55-T243	BLAST_PRODOM
					IMMUNOGLOBULIN DM00001|Q01965|139-210: M161-S232	BLAST_DOMO
					B-CELL SURFACE GLYCOPROTEIN BLAST-1 DM03635|P10252|1-239: V31-S232 DM03635|P18181|1-239: L32-S232	BLAST_DOMO
9	7682434CD1	336	S37 S175 S203 S207 S225 S282 S303 T43 T91 T143 T165 T219 T269 T290	N41 N49 N67 N137 N280 N288	signal_cleavage: M1-S30	SPSCAN
					Signal Peptides: M1-R26, M1-S30	HMMER
					Immunoglobulin domain: G231-A293, G47-F114, G147-T197	HMMER_PFAM
					PRECURSOR SIGNAL GLYCOPROTEIN IMMUNOGLOBULIN FOLD CELL ADHESION GPI-ANCHOR PROTEIN ALTERNATIVE PD005605: F35-Q124	BLAST_PRODOM

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
					IMMUNOGLOBULIN DM00001|P32736|39-125: D40-T123 DM00001|P32736|139-212: V136-D206 DM00001|P32736|226-306: I220-A302	BLAST_DOMO
10	2202389CD1	241	S44 S88 S112 S163 T10 T134 Y39	N75 N94 N110 N213	signal_cleavage: M1-I25	SPSCAN
					Immunoglobulin domain: G51-A117	HMMER_PFAM
11	7503597CD1	766	S159 S207 S215 S272 S373 S387 S454 S465 S474 S507 S551 S560 S576 S690 S703 S709 S722 T230 T301 T384 T585 T630 T713 Y48 Y307 Y396	N167 N253 N324 N498	Signal Peptides: M1-E19, M1-G21, M1-Q23, M1-L22	HMMER
					Immunoglobulin domain: G62-A129, G163-A229, G433-A501, D264-V316, G349-A400	HMMER_PFAM
					Cytosolic domain: C547-V766 Transmembrane domain: V524-F546 Non-cytosolic domain: M1-A523	TMHMMER
					GLYCOPROTEIN ANTIGEN PRECURSOR PD02327: L141-I152, T169-I190	BLIMPS_PRODOM
					IRREGULAR CHIASM CROUGHEST PROTEIN PRECURSOR IRREC TRANSMEMBRANE IMMUNOGLOBULIN FOLD GLYCOPROTEIN SIGNAL CELL ADHESION PD124347: F50-V256, V261-E315	BLAST_PRODOM

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
					IMMUNOGLOBULIN DM00001|Q08180|31-126: S51-T142	BLAST_DOMO
					Leucine_Zipper: L8-L29	MOTIFS
12	7503603CD1	T6 Y24				

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
13/3855123CB1/ 2691	1-751, 214-969, 282-871, 354-864, 370-934, 388-852, 390-848, 417-958, 459-874, 496-1309, 560-1313, 575-1092, 609-1298, 662-1207, 662-1300, 671-1317, 729-1212, 731-1313, 805-1157, 884-1300, 915-1566, 916-1566, 1010-1503, 1018-1306, 1038-1273, 1038-1641, 1070-1566, 1143-1524, 1231-1791, 1280-1566, 1499-2164, 1559-2094, 1559-2130, 1559-2199, 1559-2220, 1744-1982, 1752-2366, 1770-2621, 1785-2365, 1785-2576, 1815-2098, 1815-2355, 1818-2268, 1826-2440, 1846-2604, 1892-2548, 1898-2664, 1958-2234, 1958-2453, 1969-2691, 1976-2621, 1977-2691, 1985-2631, 1992-2564, 1993-2691, 2014-2625, 2017-2625, 2039-2625, 2081-2689, 2104-2665, 2109-2444, 2140-2691, 2175-2561, 2200-2691, 2217-2691, 2233-2677, 2309-2691, 2338-2691, 2360-2675, 2444-2691, 2505-2691
14/4547188CB1/ 2518	1-148, 1-606, 1-762, 40-275, 147-553, 533-897, 736-1036, 736-1069, 861-1483, 870-1149, 879-1056, 900-1427, 920-1229, 923-1356, 1020-1643, 1049-1251, 1049-1312, 1049-1654, 1110-1537, 1114-1678, 1145-1500, 1152-1687, 1156-1687, 1190-1458, 1190-1836, 1199-1828, 1200-1793, 1222-1768, 1251-1817, 1300-1453, 1321-1944, 1321-1966, 1344-1925, 1360-2011, 1364-2011, 1397-1864, 1420-1930, 1422-2081, 1429-2001, 1438-1864, 1470-1522, 1487-1522, 1496-2172, 1500-1552, 1516-1697, 1544-2056, 1596-2090, 1597-2204, 1599-1803, 1602-2252, 1603-2197, 1642-2219, 1669-2081, 1686-2215, 1722-2336, 1767-2432, 1786-2448, 1813-2313, 1827-2479, 1833-2266, 1850-2021, 1863-2465, 2015-2210, 2015-2497, 2015-2499, 2034-2315, 2034-2507, 2034-2518, 2045-2430, 2162-2378
15/3939883CB1/ 1522	1-274, 1-467, 71-510, 124-727, 124-753, 124-794, 124-827, 124-892, 140-599, 180-982, 264-821, 274-929, 496-799, 496-954, 517-1052, 602-1075, 613-727, 717-1177, 796-887, 801-1307, 801-1485, 975-1231, 992-1522, 995-1480, 1059-1498, 1073-1522
16/3163819CB1/	1-287, 1-496, 1-502, 1-646, 1-650, 1-660, 1-970, 67-982, 97-694, 187-1084, 470-686
17/8518269CB1/ 1463	1-511, 1-804, 17-844, 27-305, 27-511, 38-511, 43-373, 43-511, 63-759, 119-511, 146-511, 147-511, 446-1361, 476-1073, 566-1463
18/1592646CB1/ 1557	1-758, 40-1554, 182-563, 183-738, 277-841, 360-876, 360-877, 494-694, 714-980, 853-1459, 863-1394, 886-1281, 900-1178, 909-1454, 937-1489, 949-1225, 960-1203,

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
	960-1436, 964-1245, 964-1250, 990-1552, 992-1221, 992-1531, 995-1253, 996-1535, 1010-1522, 1025-1527, 1089-1381, 1091-1310, 1119-1345, 1119-1528, 1119-1546, 1138-1355, 1140-1410, 1140-1465, 1140-1553, 1150-1555, 1158-1391, 1249-1557
19/7500191CB1/ 5553	1-500, 30-501, 214-734, 216-616, 216-618, 216-733, 216-768, 216-815, 216-877, 216-901, 219-765, 226-593, 226-617, 226-698, 228-759, 278-587, 278-697, 278-698, 278-740, 282-728, 282-729, 308-857, 316-876, 317-434, 455-988, 611-1167, 611-1202, 656-1263, 671-992, 683-1210, 738-1018, 1058-1319, 1127-5470, 1171-1303, 1193-1700, 1214-1830, 1223-1830, 1255-1915, 1333-1818, 1343-1625, 1438-2117, 1439-1910, 1450-1914, 1465-1910, 1493-1630, 1509-1910, 1515-2242, 1552-2101, 1606-2156, 1670-1782, 1738-2363, 1780-2114, 1780-2313, 1814-2267, 1859-1978, 1886-2453, 1895-2489, 1910-2565, 1942-2600, 2049-2396, 2114-2693, 2243-2484, 2421-2762, 2453-2637, 2665-3303, 2722-3012, 2731-3272, 2735-2970, 2778-3352, 2798-3240, 2819-3125, 2910-3495, 2971-3570, 3001-3281, 3050-3629, 3112-3767, 3147-3428, 3147-3715, 3201-3446, 3201-3568, 3201-3603, 3201-3698, 3201-3763, 3205-3659, 3239-3473, 3280-3894, 3280-3895, 3289-3949, 3419-4085, 3464-3693, 3476-4043, 3491-3784, 3492-4014, 3506-3987, 3546-4147, 3611-4177, 3620-4185, 3628-3868, 3656-4218, 3679-3889, 3679-4107, 3681-4028, 3690-4093, 3714-4354, 3719-4240, 3726-4245, 3773-4022, 3784-4262, 3797-4056, 3798-4065, 3855-3979,
	3872-4089, 3889-4393, 3930-4543, 3947-4186, 3998-4654, 4005-4497, 4007-4201, 4017-4613, 4033-4246, 4033-4563, 4053-4417, 4053-4441, 4059-4716, 4065-4531, 4066-4372, 4073-4530, 4081-4419, 4088-4496, 4089-4678, 4148-4688, 4150-4681,
	4195-4420, 4195-4431, 4195-4478, 4195-4708, 4195-4758, 4195-4828, 4204-4515, 4219-4809, 4229-4468, 4245-4745, 4245-4888, 4252-4878, 4255-4507, 4256-4500, 4274-4491, 4280-5057, 4281-4833, 4306-4866, 4314-4846, 4330-4828, 4330-4899,
	4336-4805, 4358-4635, 4425-4880, 4426-4709, 4429-4770, 4429-4774, 4432-4952, 4444-4665, 4445-5152, 4506-5025, 4512-5096, 4590-5188, 4599-5150, 4602-5114, 4618-5153, 4651-4800, 4651-5127, 4652-5151, 4722-5146, 4736-5016, 4741-5213,
	4746-4970, 4746-5207, 4746-5213, 4747-4965, 4750-5165, 4752-5213, 4754-4981, 4754-5215, 4756-5159, 4771-5217, 4779-5213, 4785-5076, 4787-5164, 4790-5032, 4811-5092, 4812-5211, 4818-5217, 4835-5100, 4844-5158, 4844-5165,

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
	4854-5164, 4854-5217, 4863-5006, 4863-5197, 4863-5212, 4867-5163, 4867-5165, 4870-5057, 4881-5089, 4883-5138, 4885-5168, 4888-5149, 4907-5270, 4965-5219, 4966-5154, 4966-5217, 4991-5219, 4993-5164, 5050-5213, 5111-5217, 5283-5553, 5285-5525
20/7500099CB1/ 1849	1-270, 1-375, 1-400, 1-475, 1-534, 1-536, 4-632, 4-1847, 22-280, 28-293, 51-642, 123-754, 307-902, 411-1001, 434-969, 437-902, 447-1073, 450-1063, 485-1003, 487-1082, 498-988, 557-1131, 571-1069, 580-1191, 642-801, 729-1012, 810-1094,
	810-1098, 810-1102, 879-1039, 879-1391, 1100-1420, 1100-1628, 1100-1639, 1100-1647, 1100-1671, 1100-1775, 1101-1673, 1170-1782, 1241-1653,
	1251-1849, 1264-1849, 1271-1654, 1274-1704, 1345-1732, 1464-1757, 1475-1807, 1483-1738, 1615-1821
21/7682434CB1/ 1427	1-575, 72-612, 260-466, 341-923, 341-944, 341-974, 341-976, 341-1072, 371-891, 373-934, 632-1085, 676-935, 676-1093, 759-1275, 856-1156, 856-1427
22/2202389CB1/ 1014	1-365, 1-510, 246-507, 246-739, 336-990, 509-1013, 550-1014, 556-1014, 574-1014, 593-1014, 599-1014, 605-1014, 734-1007, 769-899
23/7503597CB1/ 3695	1-642, 1-803, 1-820, 26-820, 54-590, 71-625, 197-821, 528-820, 618-1141, 618-1162, 620-806, 620-820, 849-1210, 922-1166, 922-1210, 945-1723, 946-1385, 946-1524, 946-1621, 946-1643, 946-1649, 946-1687, 946-1692, 970-1466, 1001-1770, 1006-1781, 1012-1354, 1015-1433, 1046-1792, 1046-3681, 1057-1524, 1060-1526, 1060-1846, 1205-1972, 1208-1731, 1211-1845, 1440-2053, 1920-2563, 1921-2563, 2015-2508, 2043-2278, 2043-2537, 2043-2599, 2043-2646, 2043-2657, 2043-2703, 2043-2759, 2043-2792, 2043-2872, 2044-2721, 2075-2563, 2236-2796, 2285-2563, 2473-3226, 2521-3367, 2572-3099, 2572-3135, 2572-3204, 2572-3225, 2572-3232, 2621-3523, 2733-3345, 2749-2988, 2775-3626, 2823-3103, 2823-3360, 2824-3273, 2831-3445, 2853-3629, 2917-3607, 2920-3695, 2947-3695, 2951-3629, 2957-3291, 2964-3239, 2964-3458, 2967-3630, 2976-3504, 2981-3626, 2983-3692, 2984-3629, 2984-3636, 2995-3569, 2995-3693, 3010-3630, 3019-3630, 3029-3695, 3036-3425, 3044-3630, 3048-3521, 3049-3693, 3077-3695, 3109-3670, 3117-3695, 3126-3695, 3145-3676, 3180-3566, 3204-3695, 3205-3680, 3220-3695, 3222-3695, 3238-3681, 3255-3695, 3311-3695, 3314-3695, 3343-3695, 3365-3680, 3449-3695
24/7503603CB1/ 2403	1-212, 1-818, 1-829, 1-2397, 612-758, 612-815, 612-1152, 622-911, 737-1359, 746-1024, 755-930, 776-1303, 789-902, 797-1100, 897-1519, 925-1127, 925-1188, 925-1530, 990-1554, 1022-1376, 1066-1334, 1066-1712, 1075-1704, 1076-1669, 1127-1693,

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
	1197-1820, 1197-1838, 1220-1800, 1236-1889, 1259-2127, 1264-2127, 1273-1740, 1296-1805, 1299-2127, 1314-1740, 1328-2127, 1346-1398, 1350-2127,
	1376-1428, 1377-2049, 1420-1934, 1470-2324, 1472-1968, 1473-2081, 1475-1679, 1478-2130, 1479-2075, 1513-2126, 1514-2323, 1531-2326, 1545-1959,
	1562-2093, 1579-2325, 1599-2214, 1600-2326, 1623-2127, 1643-2310, 1662-2326, 1689-2191, 1703-2357, 1709-2145, 1739-2343, 1845-2295, 1869-2317, 1895-2382, 1946-2273, 2040-2256, 2104-2403

Table 5

Polynucleotide SEQ ID NO:	Incyte Project ID:	Representative Library
13	3855123CB1	BRAHNON05
14	4547188CB1	COLXTDT01
15	3939883CB1	SKINBIT01
16	3163819CB1	TLYMTXT04
17	8518269CB1	TLYJTXF01
18	1592646CB1	EOSIHET02
19	7500191CB1	BRAIFER05
20	7500099CB1	LUNGDIN02
21	7682434CB1	BRABDIK02
22	2202389CB1	SPLNFET02
23	7503597CB1	BRAHNON05
24	7503603CB1	COLXTDT01

Table 6

Library	Vector	Library Description
BRABDIK02	PSPORT1	This amplified and normalized library was constructed using pooled cDNA from three different donors. cDNA was generated using mRNA isolated from diseased vermis tissue removed from a 79-year-old Caucasian female (donor A) who died from pneumonia, an 83-year-old Caucasian male (donor B) who died from congestive heart failure, and an 87-year-old Caucasian female (donor C) who died from esophageal cancer. Pathology indicated severe Alzheimer's disease in donors A & B and moderate Alzheimer's disease in donor C. Patient history included glaucoma, pseudophakia, gastritis with gastrointestinal bleeding, peripheral vascular disease, chronic obstructive pulmonary disease, seizures, tobacco abuse in remission, and transitory ischemic attacks in donor A; Parkinson's disease and atherosclerosis in donor B; hypertension, coronary artery disease, cerebral vascular accident, and hypothyroidism in donor C. Family history included Alzheimer's disease in the mother and sibling(s) of donor A. Independent clones from this amplified library were normalized in one round using conditions adapted Soares et al., PNAS (1994) 91:9228-9232 and Bonaldo et al., Genome Research 6 (1996):791, except that a significantly longer (48 hours/round) reannealing hybridization was used.
BRAHNON05	pINCY	This normalized hippocampus tissue library was constructed from 1.6 million independent clones from a hippocampus tissue library. Starting RNA was made from posterior hippocampus removed from a 35-year-old Caucasian male who died from cardiac failure. Pathology indicated moderate leptomeningeal fibrosis and multiple microinfarctions of the cerebral neocortex. The cerebral hemisphere revealed moderate fibrosis of the leptomeninges with focal calcifications. There was evidence of shrunken and slightly eosinophilic pyramidal neurons throughout the cerebral hemispheres. There were small microscopic areas of cavitation with gliosis, scattered through the cerebral cortex. Patient history included cardiomyopathy, CHF, cardiomegaly, an enlarged spleen and liver. Patient medications included simethicone, Lasix, Digoxin, Colace, Zantac, captopril, and Vasotec. The library was normalized in two rounds using conditions adapted from Soares et al., PNAS (1994) 91:9228 and Bonaldo et al., Genome Research 6 (1996):791, except that a significantly longer (48 hours/round) reannealing hybridization was used.
BRAIFER05	pINCY	Library was constructed using RNA isolated from brain tissue removed from a Caucasian male fetus who was stillborn with a hypoplastic left heart at 23 weeks' gestation.
COLXTDT01	pINCY	Library was constructed using RNA isolated from colon tissue removed from the appendix of a 37-year-old Black female during myomectomy, dilation and curettage, right fimbrial region biopsy, and incidental appendectomy. Pathology indicated an unremarkable appendix. Pathology for the associated tumor tissue indicated multiple (12) uterine leiomyomata. Patient history included premenopausal menorrhagia and sarcoidosis of the lung. Family history included acute myocardial infarction and atherosclerotic coronary artery disease.

Table 6

Library	Vector	Library Description
EOSIHET02	PBLUESCRIPT	Library was constructed using RNA isolated from peripheral blood cells apheresed from a 48-year-old Caucasian male. Patient history included hypereosinophilia. The cell population was determined to be greater than 77% eosinophils by Wright's staining.
LUNGDIN02	pINCY	This normalized lung tissue library was constructed from 7.6x10e5 independent clones from a diseased lung tissue library. Starting RNA was made from RNA isolated from diseased lung tissue. Pathology indicated ideopathic pulmonary disease. The library was normalized in 2 rounds using conditions adapted from Soares et al., PNAS (1994) 91:9228-9232 and Bonaldo et al., Genome Research 6 (1996):791, except that a significantly longer (48 hours/round) reannealing hybridization was used.
SKINBIT01	pINCY	Library was constructed using RNA isolated from diseased skin tissue of the left lower leg. Patient history included erythema nodosum of the left lower leg.
SPLNFET02	pINCY	Library was constructed using RNA isolated from spleen tissue removed from a Caucasian male fetus, who died at 23 weeks' gestation.
TLYJTXF01	PRARE	This 5' cap isolated full-length library was constructed using RNA isolated from a treated Jurkat cell line derived from the T cells of a male. The cells were treated with 5nM of PMA and 50ng/mL of Ionomycin for 1 hour. Patient history included acute T-cell leukemia.
TLYMTXT04	pINCY	Library was constructed using RNA isolated from CD4+ T cells obtained from a pool of donors. The cells were treated with CD3 and CD28 antibodies.

Table 7

Program	Description	Reference	Parameter Threshold
ABI FACTURA	A program that removes vector sequences and masks ambiguous bases in nucleic acid sequences.	Applied Biosystems, Foster City, CA.	
ABI/PARACEL FDF	A Fast Data Finder useful in comparing and annotating amino acid or nucleic acid sequences.	Applied Biosystems, Foster City, CA; Paracel Inc., Pasadena, CA.	Mismatch <50%
ABI AutoAssembler	A program that assembles nucleic acid sequences.	Applied Biosystems, Foster City, CA.	
BLAST	A Basic Local Alignment Search Tool useful in sequence similarity search for amino acid and nucleic acid sequences. BLAST includes five functions: blastp, blastn, blastx, tblastn, and tblastx.	Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410; Altschul, S.F. et al. (1997) Nucleic Acids Res. 25:3389-3402.	ESTs: Probability value= 1.0E-8 or less; Full Length sequences: Probability value= 1.0E-10 or less
FASTA	A Pearson and Lipman algorithm that searches for similarity between a query sequence and a group of sequences of the same type. FASTA comprises as least five functions: fasta, tfasta, fastx, tfastx, and ssearch.	Pearson, W.R. and D.J. Lipman (1988) Proc. Natl. Acad Sci. USA 85:2444-2448; Pearson, W.R. (1990) Methods Enzymol. 183:63-98; and Smith, T.F. and M.S. Waterman (1981) Adv. Appl. Math. 2:482-489.	ESTs: fasta E value=1.06E-6; Assembled ESTs: fasta Identity= 95% or greater and Match length=200 bases or greater; fastx E value=1.0E-8 or less; Full Length sequences: fastx score=100 or greater
BLIMPS	A BLocks IMProved Searcher that matches a sequence against those in BLOCKS, PRINTS, DOMO, PRODOM, and PFAM databases to search for gene families, sequence homology, and structural fingerprint regions.	Henikoff, S. and J.G. Henikoff (1991) Nucleic Acids Res. 19:6565-6572; Henikoff, J.G. and S. Henikoff (1996) Methods Enzymol. 266:88-105; and Attwood, T.K. et al. (1997) J. Chem. Inf. Comput. Sci. 37:417-424.	Probability value= 1.0E-3 or less

Table 7

Program	Description	Reference	Parameter Threshold
HMMER	An algorithm for searching a query sequence against hidden Markov model (HMM)-based databases of protein family consensus sequences, such as PFAM, SMART and TIGRFAM.	Krogh, A. et al. (1994) J. Mol. Biol. 235:1501-1531; Sonnhammer, E.L.L. et al. (1988) Nucleic Acids Res. 26:320-322; Durbin, R. et al. (1998) Our World View, in a Nutshell, Cambridge Univ. Press, pp. 1-350.	PFAM, SMART or TIGRFAM hits: Probability value= 1.0E-3 or less; Signal peptide hits: Score= 0 or greater
ProfileScan	An algorithm that searches for structural and sequence motifs in protein sequences that match sequence patterns defined in Prosite.	Gribskov, M. et al. (1988) CABIOS 4:61-66; Gribskov, M. et al. (1989) Methods Enzymol. 183:146-159; Bairoch, A. et al. (1997) Nucleic Acids Res. 25:217-221.	Normalized quality score≥GCG-specified "HIGH" value for that particular Prosite motif. Generally, score=1.4-2.1.
Phred	A base-calling algorithm that examines automated sequencer traces with high sensitivity and probability.	Ewing, B. et al. (1998) Genome Res. 8:175-185; Ewing, B. and P. Green (1998) Genome Res. 8:186-194.	
Phrap	A Phils Revised Assembly Program including SWAT and CrossMatch, programs based on efficient implementation of the Smith-Waterman algorithm, useful in searching sequence homology and assembling DNA sequences.	Smith, T.F. and M.S. Waterman (1981) Adv. Appl. Math. 2:482-489; Smith, T.F. and M.S. Waterman (1981) J. Mol. Biol. 147:195-197; and Green, P., University of Washington, Seattle, WA.	Score= 120 or greater; Match length= 56 or greater
Consed	A graphical tool for viewing and editing Phrap assemblies.	Gordon, D. et al. (1998) Genome Res. 8:195-202.	
SPScan	A weight matrix analysis program that scans protein sequences for the presence of secretory signal peptides.	Nielson, H. et al. (1997) Protein Engineering 10:1-6; Claverie, J.M. and S. Audic (1997) CABIOS 12:431-439.	Score=3.5 or greater
TMAP	A program that uses weight matrices to delineate transmembrane segments on protein sequences and determine orientation.	Persson, B. and P. Argos (1994) J. Mol. Biol. 237:182-192; Persson, B. and P. Argos (1996) Protein Sci. 5:363-371.	

Table 7

Program	Description	Reference	Parameter Threshold
TMHMMER	A program that uses a hidden Markov model (HMM) to delineate transmembrane segments on protein sequences and determine orientation.	Sonnhammer, E.L. et al. (1998) Proc. Sixth Intl. Conf. On Intelligent Systems for Mol. Biol., Glasgow et al., eds., The Am. Assoc. for Artificial Intelligence (AAAI) Press, Menlo Park, CA, and MIT Press, Cambridge, MA, pp. 175-182.	
Motifs	A program that searches amino acid sequences for patterns that matched those defined in Prosite.	Bairoch, A. et al. (1997) Nucleic Acids Res. 25:217-221; Wisconsin Package Program Manual, version 9, page M51-59, Genetics Computer Group, Madison, WI.	

What is claimed is:

1. An isolated polypeptide selected from the group consisting of:

a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-12,

b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-6 and SEQ ID NO:8-12,

c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-12, and

d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-12.

2. An isolated polypeptide of claim 1 comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-12.

3. An isolated polynucleotide encoding a polypeptide of claim 1.

4. An isolated polynucleotide encoding a polypeptide of claim 2.

5. An isolated polynucleotide of claim 4 comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:13-24.

6. A recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide of claim 3.

7. A cell transformed with a recombinant polynucleotide of claim 6.

8. A transgenic organism comprising a recombinant polynucleotide of claim 6.

9. A method of producing a polypeptide of claim 1, the method comprising:

a) culturing a cell under conditions suitable for expression of the polypeptide, wherein said cell is transformed with a recombinant polynucleotide, and said recombinant

polynucleotide comprises a promoter sequence operably linked to a polynucleotide encoding the polypeptide of claim 1, and

b) recovering the polypeptide so expressed.

10. A method of claim 9, wherein the polypeptide comprises an amino acid sequence selected from the group consisting of SEQ ID NO:1-12.

11. An isolated antibody which specifically binds to a polypeptide of claim 1.

12. An isolated polynucleotide selected from the group consisting of:

a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:13-24,

b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:13-18 and SEQ ID NO:20-24,

c) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 94% identical to the polynucleotide sequence of SEQ ID NO:19,

d) a polynucleotide complementary to a polynucleotide of a),

e) a polynucleotide complementary to a polynucleotide of b),

f) a polynucleotide complementary to a polynucleotide of c), and

e) an RNA equivalent of a)-f).

13. An isolated polynucleotide comprising at least 60 contiguous nucleotides of a polynucleotide of claim 12.

14. A method of detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide of claim 12, the method comprising:

a) hybridizing the sample with a probe comprising at least 20 contiguous nucleotides comprising a sequence complementary to said target polynucleotide in the sample, and which probe specifically hybridizes to said target polynucleotide, under conditions whereby a hybridization complex is formed between said probe and said target polynucleotide or fragments thereof, and

b) detecting the presence or absence of said hybridization complex, and, optionally, if

present, the amount thereof.

15. A method of claim 14, wherein the probe comprises at least 60 contiguous nucleotides.

16. A method of detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide of claim 12, the method comprising:

a) amplifying said target polynucleotide or fragment thereof using polymerase chain reaction amplification, and

b) detecting the presence or absence of said amplified target polynucleotide or fragment thereof, and, optionally, if present, the amount thereof.

17. A composition comprising a polypeptide of claim 1 and a pharmaceutically acceptable excipient.

18. A composition of claim 17, wherein the polypeptide comprises an amino acid sequence selected from the group consisting of SEQ ID NO:1-12.

19. A method for treating a disease or condition associated with decreased expression of functional IGSFP, comprising administering to a patient in need of such treatment the composition of claim 17.

20. A method of screening a compound for effectiveness as an agonist of a polypeptide of claim 1, the method comprising:

a) exposing a sample comprising a polypeptide of claim 1 to a compound, and

b) detecting agonist activity in the sample.

21. A composition comprising an agonist compound identified by a method of claim 20 and a pharmaceutically acceptable excipient.

22. A method for treating a disease or condition associated with decreased expression of functional IGSFP, comprising administering to a patient in need of such treatment a composition of claim 21.

23. A method of screening a compound for effectiveness as an antagonist of a polypeptide of claim 1, the method comprising:

a) exposing a sample comprising a polypeptide of claim 1 to a compound, and

b) detecting antagonist activity in the sample.

24. A composition comprising an antagonist compound identified by a method of claim 23 and a pharmaceutically acceptable excipient.

25. A method for treating a disease or condition associated with overexpression of functional IGSFP, comprising administering to a patient in need of such treatment a composition of claim 24.

26. A method of screening for a compound that specifically binds to the polypeptide of claim 1, the method comprising:

a) combining the polypeptide of claim 1 with at least one test compound under suitable conditions, and

b) detecting binding of the polypeptide of claim 1 to the test compound, thereby identifying a compound that specifically binds to the polypeptide of claim 1.

27. A method of screening for a compound that modulates the activity of the polypeptide of claim 1, the method comprising:

a) combining the polypeptide of claim 1 with at least one test compound under conditions permissive for the activity of the polypeptide of claim 1,

b) assessing the activity of the polypeptide of claim 1 in the presence of the test compound, and

c) comparing the activity of the polypeptide of claim 1 in the presence of the test compound with the activity of the polypeptide of claim 1 in the absence of the test compound, wherein a change in the activity of the polypeptide of claim 1 in the presence of the test compound is indicative of a compound that modulates the activity of the polypeptide of claim 1.

28. A method of screening a compound for effectiveness in altering expression of a target polynucleotide, wherein said target polynucleotide comprises a sequence of claim 5, the method comprising:

a) exposing a sample comprising the target polynucleotide to a compound, under conditions suitable for the expression of the target polynucleotide,

b) detecting altered expression of the target polynucleotide, and

c) comparing the expression of the target polynucleotide in the presence of varying amounts of the compound and in the absence of the compound.

29. A method of assessing toxicity of a test compound, the method comprising:

a) treating a biological sample containing nucleic acids with the test compound,

b) hybridizing the nucleic acids of the treated biological sample with a probe comprising at least 20 contiguous nucleotides of a polynucleotide of claim 12 under conditions whereby a specific hybridization complex is formed between said probe and a target polynucleotide in the biological sample, said target polynucleotide comprising a polynucleotide sequence of a polynucleotide of claim 12 or fragment thereof,

c) quantifying the amount of hybridization complex, and

d) comparing the amount of hybridization complex in the treated biological sample with the amount of hybridization complex in an untreated biological sample, wherein a difference in the amount of hybridization complex in the treated biological sample is indicative of toxicity of the test compound.

30. A diagnostic test for a condition or disease associated with the expression of IGSFP in a biological sample, the method comprising:

a) combining the biological sample with an antibody of claim 11, under conditions suitable for the antibody to bind the polypeptide and form an antibody:polypeptide complex, and

b) detecting the complex, wherein the presence of the complex correlates with the presence of the polypeptide in the biological sample.

31. The antibody of claim 11, wherein the antibody is:

a) a chimeric antibody,

b) a single chain antibody,

c) a Fab fragment,

d) a $F(ab')_2$ fragment, or

e) a humanized antibody.

32. A composition comprising an antibody of claim 11 and an acceptable excipient.

33. A method of diagnosing a condition or disease associated with the expression of IGSFP in a subject, comprising administering to said subject an effective amount of the composition of claim 32.

34. A composition of claim 32, wherein the antibody is labeled.

35. A method of diagnosing a condition or disease associated with the expression of IGSFP in a subject, comprising administering to said subject an effective amount of the composition of claim 34.

36. A method of preparing a polyclonal antibody with the specificity of the antibody of claim 11, the method comprising:

a) immunizing an animal with a polypeptide consisting of an amino acid sequence selected from the group consisting of SEQ ID NO:1-12, or an immunogenic fragment thereof, under conditions to elicit an antibody response,

b) isolating antibodies from said animal, and

c) screening the isolated antibodies with the polypeptide, thereby identifying a polyclonal antibody which specifically binds to a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-12.

37. A polyclonal antibody produced by a method of claim 36.

38. A composition comprising the polyclonal antibody of claim 37 and a suitable carrier.

39. A method of making a monoclonal antibody with the specificity of the antibody of claim 11, the method comprising:

a) immunizing an animal with a polypeptide consisting of an amino acid sequence selected from the group consisting of SEQ ID NO:1-12, or an immunogenic fragment thereof, under conditions to elicit an antibody response,

b) isolating antibody producing cells from the animal,

c) fusing the antibody producing cells with immortalized cells to form monoclonal antibody-producing hybridoma cells,

d) culturing the hybridoma cells, and

e) isolating from the culture monoclonal antibody which specifically binds to a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-12.

40. A monoclonal antibody produced by a method of claim 39.

41. A composition comprising the monoclonal antibody of claim 40 and a suitable carrier.

42. The antibody of claim 11, wherein the antibody is produced by screening a Fab expression library.

43. The antibody of claim 11, wherein the antibody is produced by screening a recombinant immunoglobulin library.

44. A method of detecting a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-12 in a sample, the method comprising:

a) incubating the antibody of claim 11 with a sample under conditions to allow specific binding of the antibody and the polypeptide, and

b) detecting specific binding, wherein specific binding indicates the presence of a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-12 in the sample.

45. A method of purifying a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-12 from a sample, the method comprising:

a) incubating the antibody of claim 11 with a sample under conditions to allow specific binding of the antibody and the polypeptide, and

b) separating the antibody from the sample and obtaining the purified polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-12.

46. A microarray wherein at least one element of the microarray is a polynucleotide of claim 13.

47. A method of generating an expression profile of a sample which contains polynucleotides, the method comprising:

a) labeling the polynucleotides of the sample,

b) contacting the elements of the microarray of claim 46 with the labeled polynucleotides of the sample under conditions suitable for the formation of a hybridization complex, and

c) quantifying the expression of the polynucleotides in the sample.

48. An array comprising different nucleotide molecules affixed in distinct physical locations on a solid substrate, wherein at least one of said nucleotide molecules comprises a first oligonucleotide or polynucleotide sequence specifically hybridizable with at least 30 contiguous nucleotides of a target polynucleotide, and wherein said target polynucleotide is a polynucleotide of claim 12.

49. An array of claim 48, wherein said first oligonucleotide or polynucleotide sequence is completely complementary to at least 30 contiguous nucleotides of said target polynucleotide.

50. An array of claim 48, wherein said first oligonucleotide or polynucleotide sequence is completely complementary to at least 60 contiguous nucleotides of said target polynucleotide.

51. An array of claim 48, wherein said first oligonucleotide or polynucleotide sequence is completely complementary to said target polynucleotide.

52. An array of claim 48, which is a microarray.

53. An array of claim 48, further comprising said target polynucleotide hybridized to a nucleotide molecule comprising said first oligonucleotide or polynucleotide sequence.

54. An array of claim 48, wherein a linker joins at least one of said nucleotide molecules to said solid substrate.

55. An array of claim 48, wherein each distinct physical location on the substrate contains multiple nucleotide molecules, and the multiple nucleotide molecules at any single distinct physical location have the same sequence, and each distinct physical location on the substrate contains

nucleotide molecules having a sequence which differs from the sequence of nucleotide molecules at another distinct physical location on the substrate.

56. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:1.

57. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:2.

58. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:3.

59. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:4.

60. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:5.

61. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:6.

62. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:7.

63. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:8.

64. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:9.

65. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:10.

66. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:11.

67. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:12.

68. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:13.

69. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:14.

70. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:15.

71. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:16.

72. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:17.

73. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:18.

74. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:19.

75. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:20.

76. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:21.

77. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:22.

78. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:23.

79. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:24.

<110> INCYTE GENOMICS, INC.
YUE, Henry
XU, Yuming
THANGAVELU, Kavitha
WARREN, Bridget A.
TANG, Y. Tom
DUGGAN, Brendan M.
TRAN, Uyen K.
BAUGHN, Mariah R.
HONCHELL, Cynthia D.
BURFORD, Neil
FORSYTHE, Ian J.
YANG, Junming
MASON, Patricia M.

<120> IMMUNOLGLOBULIN SUPERFAMILY PROTEINS

<130> PF-0925 PCT

<140> To Be Assigned
<141> Herewith

<150> 60/275,249; 60/316,810; 60/323,977; 60/348,447;
60/343,880
<151> 2001-03-12; 2001-08-31; 2001-09-21; 2001-10-26;
2001-11-02

<160> 24
<170> PERL Program

<210> 1
<211> 442
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 3855123CD1

<400> 1

Met	Thr	Thr	Glu	Pro	Gln	Ser	Leu	Leu	Val	Asp	Leu	Gly	Ser	Asp	
1				5				10						15	
Ala	Ile	Phe	Ser	Cys	Ala	Trp	Thr	Gly	Asn	Pro	Ser	Leu	Thr	Ile	
				20				25						30	
Val	Trp	Met	Lys	Arg	Gly	Ser	Gly	Val	Val	Leu	Ser	Asn	Glu	Lys	
				35				40						45	
Thr	Leu	Thr	Leu	Lys	Ser	Val	Arg	Gln	Glu	Asp	Ala	Gly	Lys	Tyr	
				50				55						60	
Val	Cys	Arg	Ala	Val	Val	Pro	Arg	Val	Gly	Ala	Gly	Glu	Arg	Glu	
				65				70						75	
Val	Thr	Leu	Thr	Val	Asn	Gly	Pro	Pro	Ile	Ile	Ser	Ser	Thr	Gln	
				80				85						90	
Thr	Gln	His	Ala	Leu	His	Gly	Glu	Lys	Gly	Gln	Ile	Lys	Cys	Phe	
				95				100						105	
Ile	Arg	Ser	Thr	Pro	Pro	Pro	Asp	Arg	Ile	Ala	Trp	Ser	Trp	Lys	
				110				115						120	

```
Glu Asn Val Leu Glu Ser Gly Thr Ser Gly Arg Tyr Thr Val Glu
                125                 130                 135
Thr Ile Ser Thr Glu Glu Gly Val Ile Ser Thr Leu Thr Ile Ser
                140                 145                 150
Asn Ile Val Arg Ala Asp Phe Gln Thr Ile Tyr Asn Cys Thr Ala
                155                 160                 165
Trp Asn Ser Phe Gly Ser Asp Thr Glu Ile Ile Arg Leu Lys Glu
                170                 175                 180
Gln Gly Ser Glu Met Lys Ser Gly Ala Gly Leu Glu Ala Glu Ser
                185                 190                 195
Val Pro Met Ala Val Ile Ile Gly Val Ala Val Gly Ala Gly Val
                200                 205                 210
Ala Phe Leu Val Leu Met Ala Thr Ile Val Ala Phe Cys Cys Ala
                215                 220                 225
Arg Ser Gln Arg Asn Leu Lys Gly Val Val Ser Ala Lys Asn Asp
                230                 235                 240
Ile Arg Val Glu Ile Val His Lys Glu Pro Ala Ser Gly Arg Glu
                245                 250                 255
Gly Glu Glu His Ser Thr Ile Lys Gln Leu Met Met Asp Arg Gly
                260                 265                 270
Glu Phe Gln Gln Asp Ser Val Leu Lys Gln Leu Glu Val Leu Lys
                275                 280                 285
Glu Glu Glu Lys Glu Phe Gln Asn Leu Lys Asp Pro Thr Asn Gly
                290                 295                 300
Tyr Tyr Ser Val Asn Thr Phe Lys Glu His His Ser Thr Pro Thr
                305                 310                 315
Ile Ser Leu Ser Ser Cys Gln Pro Asp Leu Arg Pro Ala Gly Lys
                320                 325                 330
Gln Arg Val Pro Thr Gly Met Ser Phe Thr Asn Ile Tyr Ser Thr
                335                 340                 345
Leu Ser Gly Gln Gly Arg Leu Tyr Asp Tyr Gly Gln Arg Phe Val
                350                 355                 360
Leu Gly Met Gly Ser Ser Ser Ile Glu Leu Cys Glu Arg Glu Phe
                365                 370                 375
Gln Arg Gly Ser Leu Ser Asp Ser Ser Ser Phe Leu Asp Thr Gln
                380                 385                 390
Cys Asp Ser Ser Val Ser Ser Ser Gly Lys Gln Asp Gly Tyr Val
                395                 400                 405
Gln Phe Asp Lys Ala Ser Lys Ala Ser Ala Ser Ser Ser His His
                410                 415                 420
Ser Gln Ser Ser Ser Gln Asn Ser Asp Pro Ser Arg Pro Leu Gln
                425                 430                 435
Arg Arg Met Gln Thr His Val
                440

<210> 2
<211> 577
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 4547188CD1

<400> 2
Met Asp Gly Glu Ala Thr Val Lys Pro Gly Glu Gln Lys Glu Val
```

```
  1               5                  10                  15
Val Arg Arg Gly Arg Glu Val Asp Tyr Ser Arg Leu Ile Ala Gly
            20                  25                  30
Thr Leu Pro Gln Ser His Val Thr Ser Arg Arg Ala Gly Trp Lys
            35                  40                  45
Met Pro Leu Phe Leu Ile Leu Cys Leu Leu Gln Gly Ser Ser Phe
            50                  55                  60
Ala Leu Pro Gln Lys Arg Pro His Pro Arg Trp Leu Trp Glu Gly
            65                  70                  75
Ser Leu Pro Ser Arg Thr His Leu Arg Ala Met Gly Thr Leu Arg
            80                  85                  90
Pro Ser Ser Pro Leu Cys Trp Arg Glu Glu Ser Ser Phe Ala Ala
            95                  100                 105
Pro Asn Ser Leu Lys Gly Ser Arg Leu Val Ser Gly Glu Pro Gly
            110                 115                 120
Gly Ala Val Thr Ile Gln Cys His Tyr Ala Pro Ser Ser Val Asn
            125                 130                 135
Arg His Gln Arg Lys Tyr Trp Cys Cys Leu Gly Pro Pro Arg Trp
            140                 145                 150
Ile Cys Gln Thr Ile Val Ser Thr Asn Gln Tyr Thr His His Arg
            155                 160                 165
Tyr Arg Asp Arg Val Ala Leu Thr Asp Phe Pro Gln Arg Gly Leu
            170                 175                 180
Phe Val Val Arg Leu Ser Gln Leu Ser Pro Asp Asp Ile Gly Cys
            185                 190                 195
Tyr Leu Cys Gly Ile Gly Ser Glu Asn Asn Met Leu Phe Leu Ser
            200                 205                 210
Met Asn Leu Thr Ile Ser Ala Gly Pro Ala Ser Thr Leu Pro Thr
            215                 220                 225
Ala Thr Pro Ala Ala Gly Glu Leu Thr Met Arg Ser Tyr Gly Thr
            230                 235                 240
Ala Ser Pro Val Ala Asn Arg Trp Thr Pro Gly Thr Thr Gln Thr
            245                 250                 255
Leu Gly Gln Gly Thr Ala Trp Asp Thr Val Ala Ser Thr Pro Gly
            260                 265                 270
Thr Ser Lys Thr Thr Ala Ser Ala Glu Gly Arg Arg Thr Pro Gly
            275                 280                 285
Ala Thr Arg Pro Ala Ala Pro Gly Thr Gly Ser Trp Ala Glu Gly
            290                 295                 300
Ser Val Lys Ala Pro Ala Pro Ile Pro Glu Ser Pro Pro Ser Lys
            305                 310                 315
Ser Arg Ser Met Ser Asn Thr Thr Glu Gly Val Trp Glu Gly Thr
            320                 325                 330
Arg Ser Ser Val Thr Asn Arg Ala Arg Ala Ser Lys Asp Arg Arg
            335                 340                 345
Glu Met Thr Thr Thr Lys Ala Asp Arg Pro Arg Glu Asp Ile Glu
            350                 355                 360
Gly Val Arg Ile Ala Leu Asp Ala Ala Lys Lys Val Leu Gly Thr
            365                 370                 375
Ile Gly Pro Pro Ala Leu Val Ser Glu Thr Leu Ala Trp Glu Ile
            380                 385                 390
Leu Pro Gln Ala Thr Pro Val Ser Lys Gln Gln Ser Gln Gly Ser
            395                 400                 405
Ile Gly Glu Thr Thr Pro Ala Ala Gly Met Trp Thr Leu Gly Thr
            410                 415                 420
Pro Ala Ala Asp Val Trp Ile Leu Gly Thr Pro Ala Ala Asp Val
```

```
                425                 430                 435
Trp Thr Ser Met Glu Ala Ala Ser Gly Glu Gly Ser Ala Ala Gly
                440                 445                 450
Asp Leu Asp Ala Ala Thr Gly Asp Arg Gly Pro Gln Ala Thr Leu
                455                 460                 465
Ser Gln Thr Pro Ala Val Gly Pro Trp Gly Pro Pro Gly Lys Glu
                470                 475                 480
Ser Ser Val Lys Arg Thr Phe Pro Glu Asp Glu Ser Ser Ser Arg
                485                 490                 495
Thr Leu Ala Pro Val Ser Thr Met Leu Ala Leu Phe Met Leu Met
                500                 505                 510
Ala Leu Val Leu Leu Gln Arg Lys Leu Trp Arg Arg Arg Thr Ser
                515                 520                 525
Gln Glu Ala Glu Arg Val Thr Leu Ile Gln Met Thr His Phe Leu
                530                 535                 540
Glu Val Asn Pro Gln Ala Asp Gln Leu Pro His Val Glu Arg Lys
                545                 550                 555
Met Leu Gln Asp Asp Ser Leu Pro Ala Gly Ala Ser Leu Thr Ala
                560                 565                 570
Pro Glu Arg Asn Pro Gly Pro
                575
```

<210> 3
<211> 385
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 3939883CD1

<400> 3

```
Met Gln Thr Ser Ser Lys Pro Ser Asp Phe Leu Asn Leu Ala Lys
  1                 5                  10                  15
Lys Lys Arg Lys Phe Ser Glu Leu Leu Thr Thr Val Val Leu Leu
                 20                  25                  30
Cys Leu Leu Thr Pro Ser Trp Thr Ser Thr Gly Arg Met Trp Ser
                 35                  40                  45
His Leu Asn Arg Leu Leu Phe Trp Ser Ile Phe Ser Ser Val Thr
                 50                  55                  60
Cys Arg Lys Ala Val Leu Asp Cys Glu Ala Met Lys Thr Asn Glu
                 65                  70                  75
Phe Pro Ser Pro Cys Leu Asp Ser Lys Thr Lys Val Val Met Lys
                 80                  85                  90
Gly Gln Asn Val Ser Met Phe Cys Ser His Lys Asn Lys Ser Leu
                 95                 100                 105
Gln Ile Thr Tyr Ser Leu Phe Arg Arg Lys Thr His Leu Gly Thr
                110                 115                 120
Gln Asp Gly Lys Gly Glu Pro Ala Ile Phe Asn Leu Ser Ile Thr
                125                 130                 135
Glu Ala His Glu Ser Gly Pro Tyr Lys Cys Lys Ala Gln Val Thr
                140                 145                 150
Ser Cys Ser Lys Tyr Ser Arg Asp Phe Ser Phe Thr Ile Val Asp
                155                 160                 165
Pro Val Thr Ser Pro Val Leu Asn Ile Met Val Ile Gln Thr Glu
                170                 175                 180
```

```
Thr Asp Arg His Ile Thr Leu His Cys Leu Ser Val Asn Gly Ser
                185                 190                 195
Leu Pro Ile Asn Tyr Thr Phe Phe Glu Asn His Val Ala Ile Ser
                200                 205                 210
Pro Ala Ile Ser Lys Tyr Asp Arg Glu Pro Ala Glu Phe Asn Leu
                215                 220                 225
Thr Lys Lys Asn Pro Gly Glu Glu Glu Glu Tyr Arg Cys Glu Ala
                230                 235                 240
Lys Asn Arg Leu Pro Asn Tyr Ala Thr Tyr Ser His Pro Val Thr
                245                 250                 255
Met Pro Ser Thr Gly Gly Asp Ser Cys Pro Phe Cys Leu Lys Leu
                260                 265                 270
Leu Leu Pro Gly Leu Leu Leu Leu Leu Val Val Ile Ile Leu Ile
                275                 280                 285
Leu Ala Phe Trp Val Leu Pro Lys Tyr Lys Thr Arg Lys Ala Met
                290                 295                 300
Arg Asn Asn Val Pro Arg Asp Arg Gly Asp Thr Ala Met Glu Val
                305                 310                 315
Gly Ile Tyr Ala Asn Ile Leu Glu Lys Gln Ala Lys Glu Glu Ser
                320                 325                 330
Val Pro Glu Val Gly Ser Arg Pro Cys Val Ser Thr Ala Gln Asp
                335                 340                 345
Glu Ala Lys His Ser Gln Glu Leu Gln Tyr Ala Thr Pro Val Phe
                350                 355                 360
Gln Glu Val Ala Pro Arg Glu Gln Glu Ala Cys Asp Ser Tyr Lys
                365                 370                 375
Ser Gly Tyr Val Tyr Ser Glu Leu Asn Phe
                380                 385
```

<210> 4
<211> 221
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 3163819CD1

<400> 4

```
Met Leu Trp Leu Phe Gln Ser Leu Leu Phe Val Phe Cys Phe Gly
  1                 5                  10                  15
Pro Gly Gln Leu Arg Asn Ile Gln Val Thr Asn His Ser Gln Leu
                 20                  25                  30
Phe Gln Asn Met Thr Cys Glu Leu His Leu Thr Cys Ser Val Glu
                 35                  40                  45
Asp Ala Asp Asp Asn Val Ser Phe Arg Trp Glu Ala Leu Gly Asn
                 50                  55                  60
Thr Leu Ser Ser Gln Pro Asn Leu Thr Val Ser Trp Asp Pro Arg
                 65                  70                  75
Ile Ser Ser Glu Gln Asp Tyr Thr Cys Ile Ala Glu Asn Ala Val
                 80                  85                  90
Ser Asn Leu Ser Phe Ser Val Ser Ala Gln Lys Leu Cys Glu Asp
                 95                 100                 105
Val Lys Ile Gln Tyr Thr Asp Thr Lys Met Ile Leu Phe Met Val
                110                 115                 120
Ser Gly Ile Cys Ile Val Phe Gly Phe Ile Ile Leu Leu Leu Leu
```

```
                125                 130                 135
Val Leu Arg Lys Arg Arg Asp Ser Leu Ser Leu Ser Thr Gln Arg
                140                 145                 150
Thr Gln Gly Pro Ala Glu Ser Ala Arg Asn Leu Glu Tyr Val Ser
                155                 160                 165
Val Ser Pro Thr Asn Asn Thr Val Tyr Ala Ser Val Thr His Ser
                170                 175                 180
Asn Arg Glu Thr Glu Ile Trp Thr Pro Arg Glu Asn Asp Thr Ile
                185                 190                 195
Thr Ile Tyr Ser Thr Ile Asn His Ser Lys Glu Ser Lys Pro Thr
                200                 205                 210
Phe Ser Arg Ala Thr Ala Leu Asp Asn Val Val
                215                 220

<210> 5
<211> 332
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 8518269CD1

<400> 5
Met Leu Trp Leu Phe Gln Ser Leu Leu Phe Val Phe Cys Phe Gly
  1                 5                  10                  15
Pro Gly Asn Val Val Ser Gln Ser Ser Leu Thr Pro Leu Met Val
                 20                  25                  30
Asn Gly Ile Leu Gly Glu Ser Val Thr Leu Pro Leu Glu Phe Pro
                 35                  40                  45
Ala Gly Glu Lys Val Asn Phe Ile Thr Trp Leu Phe Asn Glu Thr
                 50                  55                  60
Ser Leu Ala Phe Ile Val Pro His Glu Thr Lys Ser Pro Glu Ile
                 65                  70                  75
His Val Thr Asn Pro Lys Gln Gly Lys Arg Leu Asn Phe Thr Gln
                 80                  85                  90
Ser Tyr Ser Leu Gln Leu Ser Asn Leu Lys Met Glu Asp Thr Gly
                 95                 100                 105
Ser Tyr Arg Ala Gln Ile Ser Thr Lys Thr Ser Ala Lys Leu Ser
                110                 115                 120
Ser Tyr Thr Leu Arg Ile Leu Arg Gln Leu Arg Asn Ile Gln Val
                125                 130                 135
Thr Asn His Ser Gln Leu Phe Gln Asn Met Thr Cys Glu Leu His
                140                 145                 150
Leu Thr Cys Ser Val Glu Asp Ala Asp Asp Asn Val Ser Phe Arg
                155                 160                 165
Trp Glu Ala Leu Gly Asn Thr Leu Ser Ser Gln Pro Asn Leu Thr
                170                 175                 180
Val Ser Trp Asp Pro Arg Ile Ser Ser Glu Gln Asp Tyr Thr Cys
                185                 190                 195
Ile Ala Glu Asn Ala Val Ser Asn Leu Ser Phe Ser Val Ser Ala
                200                 205                 210
Gln Lys Leu Cys Glu Asp Val Lys Ile Gln Tyr Thr Asp Thr Lys
                215                 220                 225
Met Ile Leu Phe Met Val Ser Gly Ile Cys Ile Val Phe Gly Phe
                230                 235                 240
```

```
Ile Ile Leu Leu Leu Leu Val Leu Arg Lys Arg Arg Asp Ser Leu
                245                 250                 255
Ser Leu Ser Thr Gln Arg Thr Gln Gly Pro Ala Glu Ser Ala Arg
                260                 265                 270
Asn Leu Glu Tyr Val Ser Val Ser Pro Thr Asn Asn Thr Val Tyr
                275                 280                 285
Ala Ser Val Thr His Ser Asn Arg Glu Thr Glu Ile Trp Thr Pro
                290                 295                 300
Arg Glu Asn Asp Thr Ile Thr Ile Tyr Ser Thr Ile Asn His Ser
                305                 310                 315
Lys Glu Ser Lys Pro Thr Phe Ser Arg Ala Thr Ala Leu Asp Asn
                320                 325                 330
Val Val
```

<210> 6
<211> 288
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 1592646CD1

```
<400> 6
Met Leu Pro His Phe Leu Gly Gly Glu Arg Val Arg Pro Ser Pro
  1                 5                  10                  15
Gly Ser Ser Ser Ser Gly Tyr Leu Pro Thr Met Ala Leu Val Leu
                20                  25                  30
Ile Leu Gln Leu Leu Thr Leu Trp Pro Leu Cys His Thr Asp Ile
                35                  40                  45
Thr Pro Ser Val Pro Pro Ala Ser Tyr His Pro Lys Pro Trp Leu
                50                  55                  60
Gly Ala Gln Pro Ala Thr Val Val Thr Pro Gly Val Asn Val Thr
                65                  70                  75
Leu Arg Cys Arg Ala Pro Gln Pro Ala Trp Arg Phe Gly Leu Phe
                80                  85                  90
Lys Pro Gly Glu Ile Ala Pro Leu Leu Phe Arg Asp Val Ser Ser
                95                 100                 105
Glu Leu Ala Glu Phe Phe Leu Glu Glu Val Thr Pro Ala Gln Gly
               110                 115                 120
Gly Ser Tyr Arg Cys Cys Tyr Arg Arg Pro Asp Trp Gly Pro Gly
               125                 130                 135
Val Trp Ser Gln Pro Ser Asp Val Leu Glu Leu Leu Val Thr Glu
               140                 145                 150
Glu Leu Pro Arg Pro Ser Leu Val Ala Leu Pro Gly Pro Val Val
               155                 160                 165
Gly Pro Gly Ala Asn Val Ser Leu Arg Cys Ala Gly Arg Leu Arg
               170                 175                 180
Asn Met Ser Phe Val Leu Tyr Arg Glu Gly Val Ala Ala Pro Leu
               185                 190                 195
Gln Tyr Arg His Ser Ala Gln Pro Trp Ala Asp Phe Thr Leu Leu
               200                 205                 210
Gly Ala Arg Ala Pro Gly Thr Tyr Ser Cys Tyr Tyr His Thr Pro
               215                 220                 225
Ser Ala Pro Tyr Val Leu Ser Gln Arg Ser Glu Val Leu Val Ile
```

```
                230                 235                 240
Ser Trp Glu Asp Ser Gly Ser Ser Asp Tyr Thr Arg Gly Asn Leu
                245                 250                 255
Val Arg Leu Gly Leu Ala Gly Leu Val Leu Ile Ser Leu Gly Ala
                260                 265                 270
Leu Val Thr Phe Asp Trp Arg Ser Gln Asn Arg Ala Pro Ala Gly
                275                 280                 285
Ile Arg Pro
```

<210> 7
<211> 1450
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7500191CD1

<400> 7

```
Met Ala Ala Glu Arg Gly Ala Arg Arg Leu Leu Ser Thr Pro Ser
  1               5                  10                  15
Phe Trp Leu Tyr Cys Leu Leu Leu Leu Gly Arg Arg Ala Pro Gly
                20                  25                  30
Ala Ala Ala Ala Arg Ser Gly Ser Ala Pro Gln Ser Pro Gly Ala
                35                  40                  45
Ser Ile Arg Thr Phe Thr Pro Phe Tyr Phe Leu Val Glu Pro Val
                50                  55                  60
Asp Thr Leu Ser Val Arg Gly Ser Ser Val Ile Leu Asn Cys Ser
                65                  70                  75
Ala Tyr Ser Glu Pro Ser Pro Lys Ile Glu Trp Lys Lys Asp Gly
                80                  85                  90
Thr Phe Leu Asn Leu Val Ser Asp Asp Arg Arg Gln Leu Leu Pro
                95                 100                 105
Asp Gly Ser Leu Phe Ile Ser Asn Val Val His Ser Lys His Asn
               110                 115                 120
Lys Pro Asp Glu Gly Tyr Tyr Gln Cys Val Ala Thr Val Glu Ser
               125                 130                 135
Leu Gly Thr Ile Ile Ser Arg Thr Ala Lys Leu Ile Val Ala Gly
               140                 145                 150
Leu Pro Arg Phe Thr Ser Gln Pro Glu Pro Ser Ser Val Tyr Ala
               155                 160                 165
Gly Asn Asn Ala Ile Leu Asn Cys Glu Val Asn Ala Asp Leu Val
               170                 175                 180
Pro Phe Val Arg Trp Glu Gln Asn Arg Gln Pro Leu Leu Leu Asp
               185                 190                 195
Asp Arg Val Ile Lys Leu Pro Ser Gly Met Leu Val Ile Ser Asn
               200                 205                 210
Ala Thr Glu Gly Asp Gly Gly Leu Tyr Arg Cys Val Val Glu Ser
               215                 220                 225
Gly Gly Pro Pro Lys Tyr Ser Asp Glu Val Glu Leu Lys Val Leu
               230                 235                 240
Pro Asp Pro Glu Val Ile Ser Asp Leu Val Phe Leu Lys Gln Pro
               245                 250                 255
Ser Pro Leu Val Arg Val Ile Gly Gln Asp Val Val Leu Pro Cys
               260                 265                 270
```

```
Val Ala Ser Gly Leu Pro Thr Pro Thr Ile Lys Trp Met Lys Asn
                275                 280                 285
Glu Glu Ala Leu Asp Thr Glu Ser Ser Glu Arg Leu Val Leu Leu
                290                 295                 300
Ala Gly Gly Ser Leu Glu Ile Ser Asp Val Thr Glu Asp Asp Ala
                305                 310                 315
Gly Thr Tyr Phe Cys Ile Ala Asp Asn Gly Asn Glu Thr Ile Glu
                320                 325                 330
Ala Gln Ala Glu Leu Thr Val Gln Ala Gln Pro Glu Phe Leu Lys
                335                 340                 345
Gln Pro Thr Asn Ile Tyr Ala His Glu Ser Met Asp Ile Val Phe
                350                 355                 360
Glu Cys Glu Val Thr Gly Lys Pro Thr Pro Thr Val Lys Trp Val
                365                 370                 375
Lys Asn Gly Asp Met Val Ile Pro Ser Asp Tyr Phe Lys Ile Val
                380                 385                 390
Lys Glu His Asn Leu Gln Val Leu Gly Leu Val Lys Ser Asp Glu
                395                 400                 405
Gly Phe Tyr Gln Cys Ile Ala Glu Asn Asp Val Gly Asn Ala Gln
                410                 415                 420
Ala Gly Ala Gln Leu Ile Ile Leu Glu His Ala Pro Ala Thr Thr
                425                 430                 435
Gly Pro Leu Pro Ser Ala Pro Arg Asp Val Val Ala Ser Leu Val
                440                 445                 450
Ser Thr Arg Phe Ile Lys Leu Thr Trp Arg Thr Pro Ala Ser Asp
                455                 460                 465
Pro His Gly Asp Asn Leu Thr Tyr Ser Val Phe Tyr Thr Lys Glu
                470                 475                 480
Gly Ile Ala Arg Glu Arg Val Glu Asn Thr Ser His Pro Gly Glu
                485                 490                 495
Met Gln Val Thr Ile Gln Asn Leu Met Pro Ala Thr Val Tyr Ile
                500                 505                 510
Phe Arg Val Met Ala Gln Asn Lys His Gly Ser Gly Glu Ser Ser
                515                 520                 525
Ala Pro Leu Arg Val Glu Thr Gln Pro Glu Val Gln Leu Pro Gly
                530                 535                 540
Pro Ala Pro Asn Leu Arg Ala Tyr Ala Ala Ser Pro Thr Ser Ile
                545                 550                 555
Thr Val Thr Trp Glu Thr Pro Val Ser Gly Asn Gly Glu Ile Gln
                560                 565                 570
Asn Tyr Lys Leu Tyr Tyr Met Glu Lys Gly Thr Asp Lys Glu Gln
                575                 580                 585
Asp Val Asp Val Ser Ser His Ser Tyr Thr Ile Asn Gly Leu Lys
                590                 595                 600
Lys Tyr Thr Glu Tyr Ser Phe Arg Val Val Ala Tyr Asn Lys His
                605                 610                 615
Gly Pro Gly Val Ser Thr Pro Asp Val Ala Val Arg Thr Leu Ser
                620                 625                 630
Asp Val Pro Ser Ala Ala Pro Gln Asn Leu Ser Leu Glu Val Arg
                635                 640                 645
Asn Ser Lys Ser Ile Met Ile His Trp Gln Pro Pro Ala Pro Ala
                650                 655                 660
Thr Gln Asn Gly Gln Ile Thr Gly Tyr Lys Ile Arg Tyr Arg Lys
                665                 670                 675
Ala Ser Arg Lys Ser Asp Val Thr Glu Thr Leu Val Ser Gly Thr
                680                 685                 690
```

```
Gln Leu Ser Gln Leu Ile Glu Gly Leu Asp Arg Gly Thr Glu Tyr
                695                 700                 705
Asn Phe Arg Val Ala Ala Leu Thr Ile Asn Gly Thr Gly Pro Ala
                710                 715                 720
Thr Asp Trp Leu Ser Ala Glu Thr Phe Glu Ser Asp Leu Asp Glu
                725                 730                 735
Thr Arg Val Pro Glu Val Pro Ser Ser Leu His Val Arg Pro Leu
                740                 745                 750
Val Thr Ser Ile Val Val Ser Trp Thr Pro Pro Glu Asn Gln Asn
                755                 760                 765
Ile Val Val Arg Gly Tyr Ala Ile Gly Tyr Gly Ile Gly Ser Pro
                770                 775                 780
His Ala Gln Thr Ile Lys Val Asp Tyr Lys Gln Arg Tyr Tyr Thr
                785                 790                 795
Ile Glu Asn Leu Asp Pro Ser Ser His Tyr Val Ile Thr Leu Lys
                800                 805                 810
Ala Phe Asn Asn Val Gly Glu Gly Ile Pro Leu Tyr Glu Ser Ala
                815                 820                 825
Val Thr Arg Pro His Thr Asp Thr Ser Glu Val Asp Leu Phe Val
                830                 835                 840
Ile Asn Ala Pro Tyr Thr Pro Val Pro Asp Pro Thr Pro Met Met
                845                 850                 855
Pro Pro Val Gly Val Gln Ala Ser Ile Leu Ser His Asp Thr Ile
                860                 865                 870
Arg Ile Thr Trp Ala Asp Asn Ser Leu Pro Lys His Gln Lys Ile
                875                 880                 885
Thr Asp Ser Arg Tyr Tyr Thr Val Arg Trp Lys Thr Asn Ile Pro
                890                 895                 900
Ala Asn Thr Lys Tyr Lys Asn Ala Asn Ala Thr Thr Leu Ser Tyr
                905                 910                 915
Leu Val Thr Gly Leu Lys Pro Asn Thr Leu Tyr Glu Phe Ser Val
                920                 925                 930
Met Val Thr Lys Gly Arg Arg Ser Ser Thr Trp Ser Met Thr Ala
                935                 940                 945
His Gly Thr Thr Phe Glu Leu Val Pro Thr Ser Pro Pro Lys Asp
                950                 955                 960
Val Thr Val Val Ser Lys Glu Gly Lys Pro Lys Thr Ile Ile Val
                965                 970                 975
Asn Trp Gln Pro Pro Ser Glu Ala Asn Gly Lys Ile Thr Gly Tyr
                980                 985                 990
Ile Ile Tyr Tyr Ser Thr Asp Val Asn Ala Glu Ile His Asp Trp
                995                1000                1005
Val Ile Glu Pro Val Val Gly Asn Arg Leu Thr His Gln Ile Gln
               1010                1015                1020
Glu Leu Thr Leu Asp Thr Pro Tyr Tyr Phe Lys Ile Gln Ala Arg
               1025                1030                1035
Asn Ser Lys Gly Met Gly Pro Met Ser Glu Ala Val Gln Phe Arg
               1040                1045                1050
Thr Pro Lys Ala Ser Gly Ser Gly Gly Lys Gly Ser Arg Leu Pro
               1055                1060                1065
Asp Leu Gly Ser Asp Tyr Lys Pro Pro Met Ser Gly Ser Asn Ser
               1070                1075                1080
Pro His Gly Ser Pro Thr Ser Pro Leu Asp Ser Asn Met Leu Leu
               1085                1090                1095
Val Ile Ile Val Ser Val Gly Val Ile Thr Ile Val Val Val Val
               1100                1105                1110
```

```
Ile Ile Ala Val Phe Cys Thr Arg Arg Thr Thr Ser His Gln Lys
                1115                1120                1125
Lys Lys Arg Ala Ala Cys Lys Ser Val Asn Gly Ser His Lys Tyr
                1130                1135                1140
Lys Gly Asn Ser Lys Asp Val Lys Pro Pro Asp Leu Trp Ile His
                1145                1150                1155
His Glu Arg Leu Glu Leu Lys Pro Ile Asp Lys Ser Pro Asp Pro
                1160                1165                1170
Asn Pro Ile Met Thr Asp Thr Pro Ile Pro Arg Asn Ser Gln Asp
                1175                1180                1185
Ile Thr Pro Val Asp Asn Ser Met Asp Ser Asn Ile His Gln Arg
                1190                1195                1200
Arg Asn Ser Tyr Arg Gly His Glu Ser Glu Asp Ser Met Ser Thr
                1205                1210                1215
Leu Ala Gly Arg Arg Gly Met Arg Pro Lys Met Met Met Pro Phe
                1220                1225                1230
Asp Ser Gln Pro Pro Gln Pro Val Ile Ser Ala His Pro Ile His
                1235                1240                1245
Ser Leu Asp Asn Pro His His His Phe His Ser Ser Ser Leu Ala
                1250                1255                1260
Ser Pro Ala Arg Ser His Leu Tyr His Pro Gly Ser Pro Trp Pro
                1265                1270                1275
Ile Gly Thr Ser Met Ser Leu Ser Asp Arg Ala Asn Ser Thr Glu
                1280                1285                1290
Ser Val Arg Asn Thr Pro Ser Thr Asp Thr Met Pro Ala Ser Ser
                1295                1300                1305
Ser Gln Thr Cys Cys Thr Asp His Gln Asp Pro Glu Gly Ala Thr
                1310                1315                1320
Ser Ser Ser Tyr Leu Ala Ser Ser Gln Glu Glu Asp Ser Gly Gln
                1325                1330                1335
Ser Leu Pro Thr Ala His Val Arg Pro Ser His Pro Leu Lys Ser
                1340                1345                1350
Phe Ala Val Pro Ala Ile Pro Pro Pro Gly Pro Pro Thr Tyr Asp
                1355                1360                1365
Pro Ala Leu Pro Ser Thr Pro Leu Leu Ser Gln Gln Ala Leu Asn
                1370                1375                1380
His His Ile His Ser Val Lys Thr Ala Ser Ile Gly Thr Leu Gly
                1385                1390                1395
Arg Ser Arg Pro Pro Met Pro Val Val Val Pro Ser Ala Pro Glu
                1400                1405                1410
Val Gln Glu Thr Thr Arg Met Leu Glu Asp Ser Glu Ser Ser Tyr
                1415                1420                1425
Glu Pro Asp Glu Leu Thr Lys Glu Met Ala His Leu Glu Gly Leu
                1430                1435                1440
Met Lys Asp Leu Asn Ala Ile Thr Thr Ala
                1445                1450

<210> 8
<211> 551
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7500099CD1
```

<400> 8

```
Met Val Ala Pro Lys Ser His Thr Asp Asp Trp Ala Pro Gly Pro
  1               5                  10                  15
Phe Ser Ser Lys Pro Gln Arg Ser Gln Leu Gln Ile Phe Ser Ser
                20                  25                  30
Val Leu Gln Thr Ser Leu Leu Phe Leu Leu Met Gly Leu Arg Ala
                35                  40                  45
Ser Gly Lys Asp Ser Ala Pro Thr Val Val Ser Gly Ile Leu Gly
                50                  55                  60
Gly Ser Val Thr Leu Pro Leu Asn Ile Ser Val Asp Thr Glu Ile
                65                  70                  75
Glu Asn Val Ile Trp Ile Gly Pro Lys Asn Ala Leu Ala Phe Ala
                80                  85                  90
Arg Pro Lys Glu Asn Val Thr Ile Met Val Lys Ser Tyr Leu Gly
                95                 100                 105
Arg Leu Asp Ile Thr Lys Trp Ser Tyr Ser Leu Cys Ile Ser Asn
               110                 115                 120
Leu Thr Leu Asn Asp Ala Gly Ser Tyr Lys Ala Gln Ile Asn Gln
               125                 130                 135
Arg Asn Phe Glu Val Thr Thr Glu Glu Glu Phe Thr Leu Phe Val
               140                 145                 150
Tyr Glu Gln Leu Gln Glu Pro Gln Val Thr Met Lys Ser Val Lys
               155                 160                 165
Val Ser Glu Asn Phe Ser Cys Asn Ile Thr Leu Met Cys Ser Val
               170                 175                 180
Lys Gly Ala Glu Lys Ser Val Leu Tyr Ser Trp Thr Pro Arg Glu
               185                 190                 195
Pro His Ala Ser Glu Ser Asn Gly Gly Ser Ile Leu Thr Val Ser
               200                 205                 210
Arg Thr Pro Cys Asp Pro Asp Leu Pro Tyr Ile Cys Thr Ala Gln
               215                 220                 225
Asn Pro Val Ser Gln Arg Ser Ser Leu Pro Val His Val Gly Gln
               230                 235                 240
Phe Cys Thr Asp Pro Gly Ala Ser Arg Gly Gly Thr Thr Gly Glu
               245                 250                 255
Thr Val Val Gly Val Leu Gly Glu Pro Val Thr Leu Pro Leu Ala
               260                 265                 270
Leu Pro Ala Cys Arg Asp Thr Glu Lys Val Val Trp Leu Phe Asn
               275                 280                 285
Thr Ser Ile Ile Ser Lys Glu Arg Glu Glu Ala Ala Thr Ala Asp
               290                 295                 300
Pro Leu Ile Lys Ser Arg Asp Pro Tyr Lys Asn Arg Val Trp Val
               305                 310                 315
Ser Ser Gln Asp Cys Ser Leu Lys Ile Ser Gln Leu Lys Ile Glu
               320                 325                 330
Asp Ala Gly Pro Tyr His Ala Tyr Val Cys Ser Glu Ala Ser Ser
               335                 340                 345
Val Thr Ser Met Thr His Val Thr Leu Leu Ile Tyr Arg Pro Glu
               350                 355                 360
Arg Asn Thr Lys Leu Trp Ile Gly Leu Phe Leu Met Val Cys Leu
               365                 370                 375
Leu Cys Val Gly Ile Phe Ser Trp Cys Ile Trp Lys Arg Lys Gly
               380                 385                 390
Arg Cys Ser Val Pro Ala Phe Cys Ser Ser Gln Ala Glu Ala Pro
               395                 400                 405
Ala Asp Thr Pro Gly Tyr Glu Lys Leu Asp Thr Pro Leu Arg Pro
```

```
                410                 415                 420
Ala Arg Gln Gln Pro Thr Pro Thr Ser Asp Ser Ser Ser Asp Ser
                425                 430                 435
Asn Leu Thr Thr Glu Glu Asp Glu Asp Arg Pro Glu Val His Lys
                440                 445                 450
Pro Ile Ser Gly Arg Tyr Glu Val Phe Asp Gln Val Thr Gln Glu
                455                 460                 465
Gly Ala Gly His Asp Pro Ala Pro Glu Gly Gln Ala Asp Tyr Asp
                470                 475                 480
Pro Val Thr Pro Tyr Val Thr Glu Val Glu Ser Val Val Gly Glu
                485                 490                 495
Asn Thr Met Tyr Ala Gln Val Phe Asn Leu Gln Gly Lys Thr Pro
                500                 505                 510
Val Ser Gln Lys Glu Glu Ser Ser Ala Thr Ile Tyr Cys Ser Ile
                515                 520                 525
Arg Lys Pro Gln Val Val Pro Pro Pro Gln Gln Asn Asp Leu Glu
                530                 535                 540
Ile Pro Glu Ser Pro Thr Tyr Glu Asn Phe Thr
                545                 550
```

<210> 9
<211> 336
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7682434CD1

<400> 9

```
Met Pro Pro Pro Ala Pro Gly Ala Arg Leu Arg Leu Leu Ala Ala
  1                 5                  10                  15
Ala Ala Leu Ala Gly Leu Ala Val Ile Ser Arg Gly Leu Leu Ser
                 20                  25                  30
Gln Ser Leu Glu Phe Asn Ser Pro Ala Asp Asn Tyr Thr Val Cys
                 35                  40                  45
Glu Gly Asp Asn Ala Thr Leu Ser Cys Phe Ile Asp Glu His Val
                 50                  55                  60
Thr Arg Val Ala Trp Leu Asn Arg Ser Asn Ile Leu Tyr Ala Gly
                 65                  70                  75
Asn Asp Arg Trp Thr Ser Asp Pro Arg Val Arg Leu Leu Ile Asn
                 80                  85                  90
Thr Pro Glu Glu Phe Ser Ile Leu Ile Thr Glu Val Gly Leu Gly
                 95                 100                 105
Asp Glu Gly Leu Tyr Thr Cys Ser Phe Gln Thr Arg His Gln Pro
                110                 115                 120
Tyr Thr Thr Gln Val Tyr Leu Ile Val His Val Pro Ala Arg Ile
                125                 130                 135
Val Asn Ile Ser Ser Pro Val Thr Val Asn Glu Gly Gly Asn Val
                140                 145                 150
Asn Leu Leu Cys Leu Ala Val Gly Arg Pro Glu Pro Thr Val Thr
                155                 160                 165
Trp Arg Gln Leu Arg Asp Gly Phe Thr Ser Glu Gly Glu Ile Leu
                170                 175                 180
Glu Ile Ser Asp Ile Gln Arg Gly Gln Ala Gly Glu Tyr Glu Cys
                185                 190                 195
```

```
Val Thr His Asn Gly Val Asn Ser Ala Pro Asp Ser Arg Arg Val
                200                 205                 210
Leu Val Thr Val Asn Tyr Pro Pro Thr Ile Thr Asp Val Thr Ser
                215                 220                 225
Ala Arg Thr Ala Leu Gly Arg Ala Ala Leu Leu Arg Cys Glu Ala
                230                 235                 240
Met Ala Val Pro Pro Ala Asp Phe Gln Trp Tyr Lys Asp Asp Arg
                245                 250                 255
Leu Leu Ser Ser Gly Thr Ala Glu Gly Leu Lys Val Gln Thr Glu
                260                 265                 270
Arg Thr Arg Ser Met Leu Leu Phe Ala Asn Val Ser Ala Arg His
                275                 280                 285
Tyr Gly Asn Tyr Thr Cys Arg Ala Ala Asn Arg Leu Gly Ala Ser
                290                 295                 300
Ser Ala Ser Met Arg Leu Leu Arg Pro Gly Ser Leu Glu Asn Ser
                305                 310                 315
Ala Pro Arg Pro Pro Gly Leu Leu Ala Leu Leu Ser Ala Leu Gly
                320                 325                 330
Trp Leu Trp Trp Arg Met
                335
```

<210> 10
<211> 241
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2202389CD1

<400> 10

```
Met Lys Thr Leu Pro Ala Met Leu Gly Thr Gly Lys Leu Phe Trp
  1                 5                  10                  15
Val Phe Phe Leu Ile Pro Tyr Leu Asp Ile Trp Asn Ile His Gly
                 20                  25                  30
Lys Glu Ser Cys Asp Val Gln Leu Tyr Ile Lys Arg Gln Ser Glu
                 35                  40                  45
His Ser Ile Leu Ala Gly Asp Pro Phe Glu Leu Glu Cys Pro Val
                 50                  55                  60
Lys Tyr Cys Ala Asn Arg Pro His Val Thr Trp Cys Lys Leu Asn
                 65                  70                  75
Gly Thr Thr Cys Val Lys Leu Glu Asp Arg Gln Thr Ser Trp Lys
                 80                  85                  90
Glu Glu Lys Asn Ile Ser Phe Phe Ile Leu His Phe Glu Pro Val
                 95                 100                 105
Leu Pro Asn Asp Asn Gly Ser Tyr Arg Cys Ser Ala Asn Phe Gln
                110                 115                 120
Ser Asn Leu Ile Glu Ser His Ser Thr Thr Leu Tyr Val Thr Gly
                125                 130                 135
Lys Gln Asn Glu Leu Ser Asp Thr Ala Gly Arg Glu Ile Asn Leu
                140                 145                 150
Val Asp Ala His Leu Lys Ser Glu Gln Thr Glu Ala Ser Thr Arg
                155                 160                 165
Gln Asn Ser Gln Val Leu Leu Ser Glu Thr Gly Ile Tyr Asp Asn
                170                 175                 180
Asp Pro Asp Leu Cys Phe Arg Met Gln Glu Gly Ser Glu Val Tyr
```

```
                185                 190                 195
Ser Asn Pro Cys Leu Glu Glu Asn Lys Pro Gly Ile Val Tyr Ala
                200                 205                 210
Ser Leu Asn His Ser Val Ile Gly Leu Asn Ser Arg Leu Ala Arg
                215                 220                 225
Asn Val Lys Glu Ala Pro Thr Glu Tyr Ala Ser Ile Cys Val Arg
                230                 235                 240
Ser
```

<210> 11
<211> 766
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7503597CD1

<400> 11

```
Met Lys Pro Phe Gln Leu Asp Leu Leu Phe Val Cys Phe Phe Leu
  1                 5                  10                  15
Phe Ser Gln Glu Leu Gly Leu Gln Lys Arg Gly Cys Cys Leu Val
                 20                  25                  30
Leu Gly Tyr Met Ala Lys Asp Lys Phe Arg Arg Met Asn Glu Gly
                 35                  40                  45
Gln Val Tyr Ser Phe Ser Gln Gln Pro Gln Asp Gln Val Val Val
                 50                  55                  60
Ser Gly Gln Pro Val Thr Leu Leu Cys Ala Ile Pro Glu Tyr Asp
                 65                  70                  75
Gly Phe Val Leu Trp Ile Lys Asp Gly Leu Ala Leu Gly Val Gly
                 80                  85                  90
Arg Asp Leu Ser Ser Tyr Pro Gln Tyr Leu Val Val Gly Asn His
                 95                 100                 105
Leu Ser Gly Glu His His Leu Lys Ile Leu Arg Ala Glu Leu Gln
                110                 115                 120
Asp Asp Ala Val Tyr Glu Cys Gln Ala Ile Gln Ala Ala Ile Arg
                125                 130                 135
Ser Arg Pro Ala Arg Leu Thr Val Leu Val Pro Pro Asp Asp Pro
                140                 145                 150
Val Ile Leu Gly Gly Pro Val Ile Ser Leu Arg Ala Gly Asp Pro
                155                 160                 165
Leu Asn Leu Thr Cys His Ala Asp Asn Ala Lys Pro Ala Ala Ser
                170                 175                 180
Ile Ile Trp Leu Arg Lys Gly Glu Val Ile Asn Gly Ala Thr Tyr
                185                 190                 195
Ser Lys Thr Leu Leu Arg Asp Gly Lys Arg Glu Ser Ile Val Ser
                200                 205                 210
Thr Leu Phe Ile Ser Pro Gly Asp Val Glu Asn Gly Gln Ser Ile
                215                 220                 225
Val Cys Arg Ala Thr Asn Lys Ala Ile Pro Gly Gly Lys Glu Thr
                230                 235                 240
Ser Val Thr Ile Asp Ile Gln His Pro Pro Leu Val Asn Leu Ser
                245                 250                 255
Val Glu Pro Gln Pro Val Leu Glu Asp Asn Val Val Thr Phe His
                260                 265                 270
```

```
Cys Ser Ala Lys Ala Asn Pro Ala Val Thr Gln Tyr Arg Trp Ala
                275                 280                 285
Lys Arg Gly Gln Ile Ile Lys Glu Ala Ser Gly Glu Val Tyr Arg
            290                 295                 300
Thr Thr Val Asp Tyr Thr Tyr Phe Ser Glu Pro Val Ser Cys Glu
        305                 310                 315
Val Thr Asn Ala Leu Gly Ser Thr Asn Leu Ser Arg Thr Val Asp
    320                 325                 330
Val Tyr Phe Gly Pro Arg Met Thr Thr Glu Pro Gln Ser Leu Leu
                335                 340                 345
Val Asp Leu Gly Ser Asp Ala Ile Phe Ser Cys Ala Trp Thr Gly
            350                 355                 360
Asn Pro Ser Leu Thr Ile Val Trp Met Lys Arg Gly Ser Gly Val
        365                 370                 375
Val Leu Ser Asn Glu Lys Thr Leu Thr Leu Lys Ser Val Arg Gln
    380                 385                 390
Glu Asp Ala Gly Lys Tyr Val Cys Arg Ala Val Val Pro Arg Val
                395                 400                 405
Gly Ala Gly Glu Arg Glu Val Thr Leu Thr Val Asn Gly Pro Pro
            410                 415                 420
Ile Ile Ser Ser Thr Gln Thr Gln His Ala Leu His Gly Glu Lys
        425                 430                 435
Gly Gln Ile Lys Cys Phe Ile Arg Ser Thr Pro Pro Pro Asp Arg
    440                 445                 450
Ile Ala Trp Ser Trp Lys Glu Asn Val Leu Glu Ser Gly Thr Ser
                455                 460                 465
Gly Arg Tyr Thr Val Glu Thr Ile Ser Thr Glu Glu Gly Val Ile
            470                 475                 480
Ser Thr Leu Thr Ile Ser Asn Ile Val Arg Ala Asp Phe Gln Thr
        485                 490                 495
Ile Tyr Asn Cys Thr Ala Trp Asn Ser Phe Gly Ser Asp Thr Glu
    500                 505                 510
Ile Ile Arg Leu Lys Glu Gln Glu Ser Val Pro Met Ala Val Ile
                515                 520                 525
Ile Gly Val Ala Val Gly Ala Gly Val Ala Phe Leu Val Leu Met
            530                 535                 540
Ala Thr Ile Val Ala Phe Cys Cys Ala Arg Ser Gln Arg Asn Leu
        545                 550                 555
Lys Gly Val Val Ser Ala Lys Asn Asp Ile Arg Val Glu Ile Val
    560                 565                 570
His Lys Glu Pro Ala Ser Gly Arg Glu Gly Glu Glu His Ser Thr
                575                 580                 585
Ile Lys Gln Leu Met Met Asp Arg Gly Glu Phe Gln Gln Asp Ser
            590                 595                 600
Val Leu Lys Gln Leu Glu Val Leu Lys Glu Glu Glu Lys Glu Phe
        605                 610                 615
Gln Asn Leu Lys Asp Pro Thr Asn Gly Tyr Tyr Ser Val Asn Thr
    620                 625                 630
Phe Lys Glu His His Ser Thr Pro Thr Ile Ser Leu Ser Ser Cys
                635                 640                 645
Gln Pro Asp Leu Arg Pro Ala Gly Lys Gln Arg Val Pro Thr Gly
            650                 655                 660
Met Ser Phe Thr Asn Ile Tyr Ser Thr Leu Ser Gly Gln Gly Arg
        665                 670                 675
Leu Tyr Asp Tyr Gly Gln Arg Phe Val Leu Gly Met Gly Ser Ser
    680                 685                 690
```

```
Ser Ile Glu Leu Cys Glu Arg Glu Phe Gln Arg Gly Ser Leu Ser
                695                 700                 705
Asp Ser Ser Ser Phe Leu Asp Thr Gln Cys Asp Ser Ser Val Ser
                710                 715                 720
Ser Ser Gly Lys Gln Asp Gly Tyr Val Gln Phe Asp Lys Ala Ser
                725                 730                 735
Lys Ala Ser Ala Ser Ser Ser His His Ser Gln Ser Ser Ser Gln
                740                 745                 750
Asn Ser Asp Pro Ser Arg Pro Leu Gln Arg Arg Met Gln Thr His
                755                 760                 765
Val
```

<210> 12
<211> 88
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7503603CD1

<400> 12
```
Met Asp Gly Glu Ala Thr Val Lys Pro Gly Glu Gln Lys Glu Val
  1               5                  10                  15
Val Arg Arg Gly Arg Glu Val Asp Tyr Ser Arg Leu Ile Ala Gly
                20                  25                  30
Thr Leu Pro Gln Ser His Val Leu Leu Ser Pro Phe His Lys Lys
                35                  40                  45
Asp Pro Ile Arg Asp Gly Cys Gly Arg Ala Leu Ser Pro Pro Gly
                50                  55                  60
Pro Ile Ser Gly Pro Trp Glu His Ser Gly Leu Pro Arg Pro Ser
                65                  70                  75
Ala Gly Gly Arg Arg Ala Pro Leu Gln Leu Gln Ile His
                80                  85
```

<210> 13
<211> 2691
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 3855123CB1

<400> 13
```
ctccactggt caacccttct cggtggagcc acagccaagt gctggaggac atacgtcgtc 60
actttccact gctcttgcaa aggccaaccc agctgtcacc cagtacaggt ggccaatgcg 120
gggccagatc atcaaggagg catctggaga ggtgtacagg accacagtgg actacacgta 180
cttctcagag cccgtctcct gtgaggtgac caacgcctgg gcagcaccaa cctcagccgc 240
acggttgacg tctactttgg gccccggatg accacagaac cccaatcctt gctcgtggat 300
ctgggctctg atgccatctt cagctgcgcc tggaccggca acccatccct gaccatcgtc 360
tggatgaagc ggggctccgg agtggtcctg agcaatgaga agaccctgac cctcaaatcc 420
gtgcgccagg aggacgcggg caagtacgtg tgccgggctg tggtgccccg tgtgggagcc 480
ggggagagag aggtgaccct gaccgtcaat ggacccccca tcatctccag cacccagacc 540
cagcacgccc tccacggcga gaagggccag atcaagtgct tcatccggag cacgccgccg 600
```

ccggaccgca tcgcctggtc ctggaaggag aacgttctgg agtcgggcac atcggggcgc 660
tatacggtgg agaccatcag caccgaggag ggcgtcatct ccaccctgac catcagcaac 720
atcgtgcggg ccgacttcca gaccatctac aactgcacgg cctggaacag cttcggctcc 780
gacactgaga tcatccggct caaggagcaa ggttcggaaa tgaagtcggg agccgggctg 840
gaagcagagt ctgtgccgat ggccgtcatc attggggtgg ccgtaggagc tggtgtggcc 900
ttcctcgtcc ttatggcaac catcgtggcg ttctgctgtg cccgttccca gagaaatctc 960
aaaggtgttg tgtcagccaa aaatgatatc cgagtggaaa ttgtccacaa ggaaccagcc 1020
tctggtcggg agggtgagga gcactccacc atcaagcagc tgatgatgga ccggggtgaa 1080
ttccagcaag actcagtcct gaaacagctg gaggtcctca aagaagagga gaaagagttt 1140
cagaacctga aggaccccac caatggctac tacagcgtca acaccttcaa agagcaccac 1200
tcaaccccga ccatctccct ctccagctgc cagcccgacc tgcgtcctgc gggcaagcag 1260
cgtgtgccca caggcatgtc cttcaccaac atctacagca ccctgagcgg ccagggccgc 1320
ctctacgact acgggcagcg gtttgtgctg ggcatgggca gctcgtccat cgagctttgt 1380
gagcgggagt tccagagagg ctccctcagc gacagcagct ccttcctgga cacgcagtgt 1440
gacagcagcg tcagcagcag cggcaagcag gatggctatg tgcagttcga caaggccagc 1500
aaggcttctg cttcctcctc ccaccactcc cagtcctcgt cccagaactc tgaccccagt 1560
cgacccctgc agcggcggat gcagactcac gtctaaggat cacacaccgc gggtggggac 1620
gggccaggga agaggtcagg gcacgttctg gttgtccagg gacgaggggt actttgcaga 1680
ggacaccaga attggccact tccaggacag cctcccagcg cctctgccac tgccttcctt 1740
cgaagctctg atcaagcaca aatctgggtc cccaggtgct gtgtgccaga ggtgggcggg 1800
tggggagaca gacagaggct gcggctgagt gcgctgtgct tagtgctgga cacccgtgtc 1860
cccggccctt tcctggaggc ccctctacca cctgctctgc ccacaggcac aagtggcagc 1920
tataactctg ctttcatgaa actgcggtcc actctctggt ctctctgtgg gctctacccc 1980
tcactgacca caagctctac ctacccctgt gcctgtgctc ccatacagcc ctggggagaa 2040
ggggatgacg tcttcccagc actgagctgc cccagaaacc ccggctcccc actgctgctc 2100
atagcccata ccctggaggc tgacaagcca gaaatggcct tggctaaagg agcctctctc 2160
tcaccaggct ggccgggagc ccacccccaa tttgtttggt gttttgtgtc catactcttg 2220
cagttctgtc cttggacttg atgccgctga actctgcggt gggaccggtc ccgtcagagc 2280
ctggtgtact ggggggaggg agggaggagg gagcctgtgc tgacggagca cctcgccggg 2340
tgtgcccctc ctgggctgtg tgaccccagc ctccccaccc acctcctgct ttgtgtactc 2400
ctcccctccc cctcagcaca atcggagttc atataagaag tgcgggagct tctctggtca 2460
gggttctctg aacacttatg gagagagtgc ttcctgggaa gtgtggcgtt tgaaggggct 2520
ggagggcagg tctttaagat ggcgagactg cccttctcag ctgataaaca caagaacggc 2580
gatcctgtct tcagtaaggc tccacgagaa gagaggaagt atatctacac ctcaaccctc 2640
ctagtcacca cctgaaataa atgttaggga cactacaaaa aaaaaaaaaa a 2691

<210> 14
<211> 2518
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 4547188CB1

<400> 14
ggaaggatat ggatcaatgt tttctttttt gaagctactg ttaccactcc tggaaaagtt 60
cttcaggaat aagtgacagt aagaatgaca agggattagg actggcttcc tcttataaat 120
aataaaatcc aaagagaagt gacttgagtc tccaggttta aagaagagca actagaagtc 180
gtccaaacac ctgcatctca taaggagaag aaaagtccac ctggatcttg tttctggact 240
gagatggatg gagaggccac agtgaagcct ggagaacaaa aggaagtggt gaggagagga 300
agagaagtgg actactccag gctcattgct ggcactttac cacaatctca cgtcaccagc 360
aggagggcag gatggaaaat gcccctcttc ctcatactgt gcctgctaca aggttcttct 420
ttcgcccttc cacaaaaaag accccatccg agatggctgt gggagggctc tctccctctc 480
aggacccatc tccgggccat gggaacactc aggccttcct cgcccctctg ctggcgggag 540
gagagctcct ttgcagctcc aaattcattg aagggctcaa ggctggtgtc aggggagcct 600

ggaggagctg tcaccatcca gtgccattat gccccctcat ctgtcaacag gcaccagagg 660
aagtactggt gctgtctggg gcccccaaga tggatctgcc agaccattgt gtccaccaac 720
cagtatactc accatcgcta tcgtgaccgt gtggccctca cagactttcc acagagaggc 780
ttgtttgtgg tgaggctgtc ccaactgtcc ccggatgaca tcggatgcta cctctgcggc 840
attggaagtg aaaacaacat gctgttctta agcatgaatc tgaccatctc tgcaggtccc 900
gccagcaccc tccccacagc cactccagct gctggggagc tcaccatgag atcctatgga 960
acagcgtctc cagtggccaa cagatggacc ccaggaacca cccagacctt aggacagggg 1020
acagcatggg acacagttgc ttccactcca ggaaccagca agactacagc ttcagctgag 1080
ggaagacgaa ccccaggagc aaccaggcca gcagctccag ggacaggcag ctgggcagag 1140
ggttctgtca aagcacctgc tccgattcca gagagtccac cttcaaagag cagaagcatg 1200
tccaatacaa cagaaggtgt ttgggagggc accagaagct cggtgacaaa cagggctaga 1260
gccagcaagg acaggaggga gatgacaact accaaggctg ataggccaag ggaggacata 1320
gaggggtca ggatagctct tgatgcagcc aaaaaggtcc taggaaccat tgggccacca 1380
gctctggtct cagaaacttt ggcctgggaa atcctcccac aagcaacgcc agtttctaag 1440
caacaatctc agggttccat tggagaaaca actccagctg caggcatgtg gaccttggga 1500
actccagctg cagatgtgtg gatcttggga actccagctg cagatgtgtg gaccagcatg 1560
gaggcagcat ctggggaagg aagcgctgca ggggacctag atgctgccac tggagacaga 1620
ggtccccaag caacactgag ccagaccccg gcagtaggac cctggggacc ccctggcaag 1680
gagtcctccg tgaagcgtac ttttccagaa gatgaaagca gctctcggac cctggctcct 1740
gtctctacca tgctggccct gtttatgctt atggctctgg ttctattgca aaggaagctc 1800
tggagaagga ggacctctca ggaggcagaa agggtcacct taattcagat gacacatttt 1860
ctggaagtga acccccaagc agaccagctg ccccatgtgg aaagaaagat gctccaggat 1920
gactctcttc ctgctggggc cagcctgact gccccagaga gaaatccagg accctgaggg 1980
acagagagat gaactgctca gttaccatgg gagaaggacc aagatcaaag gccttcagga 2040
ccccagcctc tttccatcat ccttcctcca cctgtgggaa gagaagctga tgcagccggt 2100
gctccaccca tggaagaaag gctggctgtc cttgggccca agaaagtcaa gcattatcca 2160
cgtccaaagg tgacaagatg actcaaagga gacttcaaga acagtgtatg aaacactgga 2220
agaggtcacc taggaaaagc atgaaatttc cattcctgaa tgtttgcaaa tagaagaggc 2280
ttccaatcag tgtggaaagt gacaaatccc ctatcaacac tcccagccct tgctgggggc 2340
tccttttctg actactgtta gcactcagcc tcccattcac atgtattata tttaagtgta 2400
ccagccttgc cttctcaagt agattctaag ctcctttaag gcagtaattg cattttatct 2460
gtctcatgat gcccccagag aacttccaac tcagtagacc ccaataatac ctgtgtgc 2518

<210> 15
<211> 1522
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 3939883CB1

<400> 15
aaaccagtat tatgcaaacc tcatccaaac cctctgattt ccttaacttg gctaagaaaa 60
agaggaagtt ctccgagtta ctcaccactg tggttctact atgccttctg accccgtctt 120
ggacttcaac tgggagaatg tggagccatt tgaacaggct cctcttctgg agcatatttt 180
cttctgtcac ttgtagaaaa gctgtattgg attgtgaggc aatgaaaaca aatgaattcc 240
cttctccatg tttggactca aagactaagg tggttatgaa gggtcaaaat gtatctatgt 300
tttgttccca taagaacaaa tcactgcaga tcacctattc attgtttcga cgtaagacac 360
acctgggaac ccaggatgga aaaggtgaac ctgcgatttt taacctaagc atcacagaag 420
cccatgaatc aggcccctac aaatgcaaag cccaagttac cagctgttca aaatacagtc 480
gtgacttcag cttcacgatt gtcgacccgg tgacttcccc agtgctgaac attatggtca 540
ttcaaacaga aacagaccga catataacat tacattgcct ctcagtcaat ggctcgctgc 600
ccatcaatta cactttcttt gaaaaccatg ttgccatatc accagctatt tccaagtatg 660
acagggagcc tgctgaattt aacttaacca agaagaatcc tggagaagag gaagagtata 720
ggtgtgaagc taaaaacaga ttgcctaact atgcaacata cagtcaccct gtcaccatgc 780

cctcaacagg cggagacagc tgtcctttct gtctgaagct actacttcca gggttattac 840
tgttgctggt ggtgataatc ctaattctgg ctttttgggt actgcccaaa tacaaaacaa 900
gaaaagctat gagaaataat gtgcccaggg accgtggaga cacagccatg gaagttggaa 960
tctatgcaaa tatccttgaa aaacaagcaa aggaggaatc tgtgccagaa gtgggatcca 1020
ggccgtgtgt ttccacagcc caagatgagg ccaaacactc ccaggagcta cagtatgcca 1080
cccccgtgtt ccaggaggtg gcaccaagag agcaagaagc ctgtgattct tataaatctg 1140
gatatgtcta ttctgaactc aacttctgaa atttacagaa acaaactaca tctcaggatg 1200
gagtctcact ctgttgccca ggctggagtt cagtggcgcg atcttggctc acttcaatct 1260
ccatcttccc agttcaagcg attctcatgc ctcgacctcc cgagtagctg ggattgcagg 1320
tgcccgctac cacgcccagc taatttttgt atttttagta gagatggggt ttcactatgg 1380
tggccaggct ggtcttgaac tcctgacctc agatgatctg cctgcctcgg cctcccaaag 1440
tgctggaact acaggcctga gccaccgtgc ccggccctga atcgctttag taagtaaagg 1500
gtctccaaga ataaaaaaaa aa 1522

<210> 16
<211> 1084
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 3163819CB1

<400> 16
ggaaagcatg ttgtggctgt tccaatcgct cctgtttgtc ttctgctttg gcccaggaca 60
actgaggaac atacaagtta ccaatcacag tcagctattt cagaatatga cctgtgagct 120
ccatctgact tgctctgtgg aggatgcaga tgacaatgtc tcattcagat gggaggcctt 180
gggaaacaca ctttcaagtc agccaaacct cactgtctcc tgggacccca ggatttccag 240
tgaacaggac tacacctgca tagcagagaa tgctgtcagt aatttatcct tctctgtctc 300
tgcccagaag ctttgcgaag atgttaaaat tcaatataca gataccaaaa tgattctgtt 360
tatggtttct gggatatgca tagtcttcgg tttcatcata ctgctgttac ttgttttgag 420
gaaaagaaga gattccctat ctttgtctac tcagcgaaca cagggccccg cagagtccgc 480
aaggaaccta gagtatgttt cagtgtctcc aacgaacaac actgtgtatg cttcagtcac 540
tcattcaaac agggaaacag aaatctggac acctagagaa aatgatacta tcacaattta 600
ctccacaatt aatcattcca aagagagtaa acccactttt tccagggcaa ctgcccttga 660
caatgtcgtg taagttgctg aaaggcctca gaggaattcg ggaatgacac gtcttctgat 720
cccatgagac agaacaaaga acaggaagct tggttcctgt tgttcctggc aacagaattt 780
gaatatctag gataggatga tcacctccag tccttcggac ttaaacctgc ctacctgagt 840
caaacaccta aggataacat catttccagc atgtggttca aataatattt tccaatccac 900
ttcaggccaa aacatgctaa agataacaca ccagcacatt gactctctct ttgataacta 960
agcaaatgga attatggttg acagagagtt tatgatccag aagacaacca cttctctcct 1020
tttagaaagc agcaggattg acttattgag aaataatgca gtgtgttggt tacatgtgta 1080
gtct 1084

<210> 17
<211> 1463
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 8518269CB1

<400> 17
caaaaacatt gactgcctca aggtctcaag caccagtctt caccgcggaa agcatgttgt 60
ggctgttcca atcgctcctg tttgtcttct gctttggccc agggaatgta gtttcacaaa 120

gcagcttaac cccattgatg gtgaacggga ttctggggga gtcagtaact cttcccctgg 180
agtttcctgc aggagagaag gtcaacttca tcacttggct tttcaatgaa acatctcttg 240
ccttcatagt accccatgaa accaaaagtc cagaaatcca cgtgactaat ccgaaacagg 300
gaaagcgact gaacttcacc cagtcctact ccctgcaact cagcaacctg aagatggaag 360
acacaggctc ttacagagcc cagatatcca caaagacctc tgcaaagctg tccagttaca 420
ctctgaggat attaagacaa ctgaggaaca tacaagttac caatcacagt cagctatttc 480
agaatatgac ctgtgagctc catctgactt gctctgtgga ggatgcagat gacaatgtct 540
cattcagatg ggaggccttg ggaaacacac tttcaagtca gccaaacctc actgtctcct 600
gggaccccag gatttccagt gaacaggact acacctgcat agcagagaat gctgtcagta 660
atttatcctt ctctgtctct gcccagaagc tttgcgaaga tgttaaaatt caatatacag 720
ataccaaaat gattctgttt atggtttctg ggatatgcat agtcttcggt ttcatcatac 780
tgctgttact tgttttgagg aaaagaagag attccctatc tttgtctact cagcgaacac 840
agggccccgc agagtccgca aggaacctag agtatgtttc agtgtctcca acgaacaaca 900
ctgtgtatgc ttcagtcact cattcaaaca gggaaacaga aatctggaca cctagagaaa 960
atgatactat cacaatttac tccacaatta atcattccaa agagagtaaa cccacttttt 1020
ccagggcaac tgcccttgac aatgtcgtgt aagttgctga aaggcctcag aggaattcgg 1080
gaatgacacg tcttctgatc ccatgagaca gaacaaagaa caggaagctt ggttcctgtt 1140
gttcctggca acagaatttg aatatctagg ataggatgat cacctccagt ccttcggact 1200
taaacctgcc tacctgagtc aaacacctaa ggataacatc atttccagca tgtggttcaa 1260
ataatatttt ccaatccact tcaggccaaa acatgctaaa gataacacac cagcacattg 1320
actctctctt tgataactaa gcaaatggaa ttatggttga cagagagttt atgatccaga 1380
agacaaccac ttctctcctt ttagaaagca gcaggattga cttattgaga aataatgcag 1440
tgtgttggtt acatgtgtag tct 1463

<210> 18
<211> 1557
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 1592646CB1

<400> 18
agcggggcac tcgcgcagaa caaagatgga gccgtggagt gccatagggc tatgacacag 60
tcccccacag gcccccacct cgatactgtc ttccgtaaat gaggatctgg gtctggtttt 120
ctgatgttgc ctcatttcct gggaggggag agggtgcgac caagccctgg ctccagctct 180
agcgggtatc tgcccaccat ggccctggtg ctgatcctcc agctgctgac cctctggcct 240
ctgtgtcaca cagacatcac tccgtctgtc cccccagctt cataccaccc taagccatgg 300
ctgggagctc agccggctac agttgtgacc cctggggtca acgtgacctt gagatgccgg 360
gcaccccaac ccgcttggag atttggactt ttcaagcctg gagagatcgc tccccttctc 420
ttccgggatg tgtcctccga gctggcagaa ttctttctgg aggaggtgac tccagcccaa 480
gggggaagtt accgctgctg ctaccgaagg ccagactggg ggccgggtgt ctggtcccag 540
cccagcgatg tcctggagct gctggtgaca gaggagctgc cgcggccgtc gctggtggcg 600
ctgcccgggc cggtggtggg tcctggcgcc aacgtgagcc tgcgctgcgc gggccgcctg 660
cggaacatga gcttcgtgct gtaccgcgag ggcgtggcgg ccccgctgca gtaccgccac 720
tccgcgcagc cctgggccga cttcacgctg ctgggcgccc gcgcccccgg cacctacagc 780
tgctactatc acacgccctc cgcgccctac gtgctgtcgc agcgcagcga ggtgctggtc 840
atcagctggg aagactctgg ctcctccgac tacacccggg ggaacctagt ccgcctgggg 900
ctggccgggc tggtcctcat ctccctgggc gcgctggtca cttttgactg gcgcagtcag 960
aaccgcgctc ctgctggtat ccgcccctga gccccaggag cactgcagcc cgagacttcc 1020
aacctgagtg gcggagaagc tgggaccctg ggctggactg tcctttcctg cagccccaca 1080
gtcctgctgg ctgagctccg cggaacggtc cttagacccc gctgtgccct gtgctgtagc 1140
ttctttccag gcctttccca aggagtagct gaaaggaaga cgcgattagt ggttaagact 1200
tccaagccag aagacagagg gttcgaatcc cagcactgcc gtctactcac tgtagtagta 1260
gcagctacag aaaggtagta gtgagacgtg aagccagctg gacttcctgg gttgaatggg 1320

gacctggaga acttttctgt cttacaagag gattgtaaaa tggaccaatc agcactctgt 1380
aagatggacc aatcagcgct ctgtaaaatg gaccaatcag caggacatgg gcgggacaa 1440
taagggaata aaagctggcg agcgcggcac cccaccagag tctgcttcca cgctgtggga 1500
gctttgttct cttgctctac acaataaatc ttgctgctgc taaaaaaaaa aaaaagg 1557

<210> 19
<211> 5553
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7500191CB1

<400> 19
tgcggccgcg ggagccgagc ttgcagcgag ggaccggctg aggcgcgcgg gagggaagga 60
ggcaagggct ccgcggcgct gtcgccgccg ctgccgctca ctctcgggga agagatggcg 120
gcggagcggg gagcccggcg actcctcagc acccctcct tctggctcta ctgcctgctg 180
ctgctcgggc gccgggcgcc gggcgccgcg gccgccagga gcggctccgc gccgcagtcc 240
ccaggagcca gcattcgaac gttcactcca ttttatttc tggtggagcc ggtggataca 300
ctctcagtta gaggctcttc tgttatatta aactgttcag catattctga gccttctcca 360
aaaattgaat ggaaaaaaga tggaactttt ttaaacttag tatcagatga tcgacgccag 420
cttctcccgg atggatcttt atttatcagc aatgtggtgc attccaaaca caataaacct 480
gatgaaggtt attatcagtg tgtggccact gttgagagtc ttggaactat tatcagtaga 540
acagcgaagc tcatagtagc aggtcttcca agatttacca gccaaccaga accttcctca 600
gtttatgctg ggaacaatgc aattctgaat tgtgaagtta atgcagattt ggtcccattt 660
gtgaggtggg aacagaacag acaacccctt cttctggatg atagagttat caaacttcca 720
agtggaatgc tggttatcag caatgcaact gaaggagatg gcgggcttta tcgctgcgta 780
gtggaaagtg gtgggccacc aaagtatagt gatgaagttg aattgaaggt tcttccagat 840
cctgaggtga tatcagactt ggtatttttg aaacagcctt ctcccttagt cagagtcatt 900
ggtcaggatg tagtgttgcc atgtgttgct tcaggacttc ctactccaac cattaaatgg 960
atgaaaaatg aggaggcact tgacacagaa agctctgaaa gattggtatt gctggcaggt 1020
ggtagcctgg agatcagtga tgttactgag gatgatgctg ggacttattt ttgtatagct 1080
gataatggaa atgagacaat tgaagctcaa gcagagctta cagtgcaagc tcaacctgaa 1140
ttcctgaagc agcctactaa tatatatgct cacgaatcta tggatattgt atttgaatgt 1200
gaagtgactg gaaaaccaac tccaactgtg aagtgggtca aaaatgggga tatggttatc 1260
ccaagtgatt attttaagat tgtaaaggaa cataatcttc aagttttggg tctggtgaaa 1320
tcagatgaag ggttctatca gtgcattgct gaaaatgatg ttggaaatgc acaagctgga 1380
gcccaactga taatccttga acatgcacca gccacaacgg gaccactgcc ttcagctcct 1440
cgggatgtcg tggcctccct ggtctctacc cgcttcatca aattgacgtg gcggacacct 1500
gcatcagatc ctcacggaga caaccttacc tactctgtgt tctacaccaa ggaagggatt 1560
gctagggaac gtgttgagaa taccagtcac ccaggagaga tgcaagtaac cattcaaaac 1620
ctaatgccag cgaccgtgta catctttaga gttatggctc aaaataagca tggctcagga 1680
gagagttcag ctccactgcg agtagaaaca caacctgagg ttcagctccc tggcccagca 1740
cctaaccttc gtgcatatgc agcttcgcct acctccatca ctgttacgtg ggaaacacca 1800
gtgtctggca atggggaaat tcagaattat aaattgtact acatggaaaa ggggactgat 1860
aaagaacagg atgttgatgt ttcaagtcac tcttacacca ttaatgggtt gaaaaaatat 1920
acagagtata gtttccgagt ggtggcctac aataaacatg gtcctggagt ttccacacca 1980
gatgttgctg ttcgaacatt gtcagatgtt cccagtgctg ctcctcagaa tctgtccttg 2040
gaagtgagaa attcaaagag tattatgatt cactggcagc cacctgctcc agccacacaa 2100
aatgggcaga ttactggcta caagattcgc taccgaaagg cctcccgaaa gagtgatgtc 2160
actgagacct tggtaagcgg gacacagctg tctcagctga ttgaaggtct tgatcggggg 2220
actgagtata atttccgagt ggctgctcta acaatcaatg gtacaggccc ggcaactgac 2280
tggctgtctg ctgaaacttt tgaaagtgac ctagatgaaa ctcgtgttcc tgaagtgcct 2340
agctctcttc acgtacgccc gctcgttact agcatcgtag tgagctggac tcctccagag 2400
aatcagaaca ttgtggtcag aggttacgcc attggttatg gcattggcag ccctcatgcc 2460

cagaccatca aagtggacta taaacagcgc tattacacca ttgaaaatct ggatcccagc 2520
tctcactatg tgattaccct gaaagcattt aataacgtgg gtgaaggcat ccccctgtat 2580
gagagtgctg tgaccaggcc tcacacagac acttctgaag ttgatttatt tgttattaat 2640
gctccataca ctccagtgcc agatcccact cccatgatgc caccagtggg agttcaggct 2700
tccattctga gtcatgacac catcaggatt acgtgggcag acaactcgct gcccaagcac 2760
cagaagatta cagactcccg atactacacc gtccgatgga aaaccaacat cccagcaaac 2820
accaagtaca agaatgcaaa tgcaaccact ttgagttatt tggtgactgg tttaaagccg 2880
aatacactct atgaattctc tgtgatggtg accaaaggtc gaagatcaag tacatggagt 2940
atgacagccc atgggaccac ctttgaatta gttccgactt ctccacccaa ggatgtgact 3000
gttgtgagta aagaggggaa acctaagacc ataattgtga attggcagcc tccctctgaa 3060
gccaatggca aaattacagg ttacatcata tattacagta cagatgtgaa tgcagagata 3120
catgactggg ttattgagcc tgttgtggga aacagactga ctcaccagat acaagagtta 3180
actcttgaca caccatacta cttcaaaatc caggcacgga actcaaaggg catgggaccc 3240
atgtctgaag ctgtccaatt cagaacacct aaagcctcag ggtctggagg gaaaggaagc 3300
cggctgccag acctaggatc cgactacaaa cctccaatga gcggcagtaa cagccctcat 3360
gggagcccca cctctcctct ggacagtaat atgctgctgg tcataattgt ttctgttggc 3420
gtcatcacca tcgtggtggt tgtgattatc gctgtctttt gtacccgtcg taccacctct 3480
caccagaaaa agaaacgagc tgcctgcaaa tcagtgaatg gctctcataa gtacaaaggg 3540
aattccaaag atgtgaaacc tccagatctc tggatccatc atgagagact ggagctgaaa 3600
cccattgata agtctccaga cccaaacccc atcatgactg atactccaat tcctcgcaac 3660
tctcaagata tcacaccagt tgacaactcc atggacagca atatccatca aaggcgaaat 3720
tcatacagag ggcatgagtc agaggacagc atgtctacac tggctggaag gcgaggaatg 3780
agaccaaaaa tgatgatgcc ctttgactcc cagccacccc agcctgtgat tagtgcccat 3840
cccatccatt ccctcgataa ccctcaccat catttccact ccagcagcct cgcttctcca 3900
gctcgcagtc atctctacca cccgggcagc ccatggccca ttggcacatc catgtccctt 3960
tcagacaggg ccaattccac agaatccgtt cgaaataccc ccagcactga caccatgcca 4020
gcctcttcgt ctcaaacatg ctgcactgat caccaggacc ctgaaggtgc taccagctcc 4080
tcttacttgg ccagctccca agaggaagat tcaggccaga gtcttcccac tgcccatgtt 4140
cgcccttccc acccattgaa gagcttcgcc gtgccagcaa tcccgcctcc aggacctccc 4200
acctatgatc ctgcattgcc aagcacacca ttactgtccc agcaagctct gaaccatcac 4260
attcactcag tgaagacagc ctccatcggg actctaggaa ggagccggcc tcctatgcca 4320
gtggttgttc ccagtgcccc tgaagtgcag gagaccacaa ggatgttgga agactccgag 4380
agtagctatg aaccagatga gctgaccaaa gagatggccc acctggaagg actaatgaag 4440
gacctaaacg ctatcacaac agcatgacga ccttcaccag gacctgactt caaacctgag 4500
tctggaagtc ttggaactta acccttgaaa acaaggaatt gtacagagta cgagaggaca 4560
gcacttgaga acacagaatg agccagcaga ctggccagcg cctctgtgta gggctggctc 4620
caggcatggc cacctgcctt cccctggtca gcctggaaga agcctgtgtc gaggcagctt 4680
ccctttgcct gctgatattc tgcaggactg ggcaccatgg gccaaaattt tgtgtccagg 4740
gaagaggcga gaagtgcaac ctgcatttca ctttgtggtc aggccgtgtc tttgtgctgt 4800
gactgcatca cctttatgga gtgtagacat tggcatttat gtacaatttt atttgtgtct 4860
tattttattt taccttcaaa aacaaaaacg ccatccaaaa ccaaggaagt ccttggtgtt 4920
ctccacaagt ggttgacatt tgactgcttg ttccaattat gtatggaaag tctttgacag 4980
tgtgggtcgt tcctggggtt ggcttgtttt ttggtttcat ttttatttct taattctgag 5040
tcattgcatc ctctaccagc tgttaatcca tcactctgag ggggaggaaa tgttgcattg 5100
ctgtttgtaa gcttttttta ttatttttt attataatta ttaaaggcct gactctttcc 5160
tctcatcact gtgagattac agatctattt gaattgaatg aaatgtaaca ttgaaaagac 5220
ttgtttgttg ctttctgtgc agtttcagta ttggggcggg tgggggggctg ggggttggta 5280
ataggaaatg gaggggctgc tgaggtcctg tgaatgtttc tgtcattgta ctttcttcca 5340
gaagcctgca gagaatggaa gcatcttctt tattgtcctt tcctggcatg tccatcctta 5400
ttgtcactac gttgcaactg gagtttgatt tggatctggt tttaaaattc ttctgtgcaa 5460
tagatgggtt tgaggattta gcggccctga tgtcttggtc atagcctggt aagaatgtcc 5520
atgctgagga gccacatgtt gtatttctaa ctg 5553

<210> 20
<211> 1849
<212> DNA

<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7500099CB1

<400> 20
aatagatcat catggtggca ccaaagagtc acacagatga ctgggctcct gggcctttct 60
ccagtaagcc acagaggagt cagctgcaaa tattctcttc tgttctacag acctctctcc 120
tcttcctgct catgggacta agagcctctg gaaaggactc agccccaaca gtggtgtcag 180
ggatcctagg gggttccgtg actctccccc taaacatctc agtagacaca gagattgaga 240
acgtcatctg gattggtccc aaaaatgctc ttgctttcgc acgtcccaaa gaaaatgtaa 300
ccattatggt caaaagctac ctgggccgac tagacatcac caagtggagt tactccctgt 360
gcatcagcaa tctgactctg aatgatgcag gatcctacaa agcccagata aaccaaagga 420
attttgaagt caccactgag gaggaattca ccctgttcgt ctatgagcag ctgcaggagc 480
cccaagtcac catgaagtct gtgaaggtgt ctgagaactt ctcctgtaac atcactctaa 540
tgtgctccgt gaaggggtca gagaaaagtg ttctgtacag ctggacccca agggaacccc 600
atgcttctga gtccaatgga ggctccattc ttaccgtctc ccgaacacca tgtgacccag 660
acctgccata catctgcaca gcccagaacc ccgtcagcca gagaagctcc ctccctgtcc 720
atgttgggca gttctgtaca gatccaggag cctccagagg aggaacaacg ggggagactg 780
tggtaggggt cctgggagag ccagtcaccc tgccacttgc actcccagcc tgccgggaca 840
cagagaaggt tgtctggttg tttaacacat ccatcattag caaagagagg gaagaagcag 900
caacggcaga tccactcatt aaatccaggg atccttacaa gaacagggtg tgggtctcca 960
gccaggactg ctccctgaag atcagccagc tgaagataga ggacgccggc ccctaccatg 1020
cctacgtgtg ctcagaggcc tccagcgtca ccagcatgac acatgtcacc ctgctcatct 1080
accgacctga gagaaacaca aagctttgga ttgggttgtt cctgatggtt tgccttctgt 1140
gcgttgggat cttcagctgg tgcatttgga agcgaaaagg acggtgttca gtcccagcct 1200
tctgttccag ccaagctgag gccccagcgg atacaccagg atatgagaag ctggacactc 1260
ccctcaggcc tgccaggcaa cagcctacac ccacctcaga cagcagctct gacagcaacc 1320
tcacaactga ggaggatgag gacaggcctg aggtgcacaa gcccatcagt ggaagatatg 1380
aggtatttga ccaggtcact caggagggcg ctggacatga cccagcccct gagggccaag 1440
cagactatga tcccgtcact ccatatgtca cggaagttga gtctgtggtt ggagagaaca 1500
ccatgtatgc acaagtgttc aacttacagg gaaagacccc agtttctcag aaggaagaga 1560
gctcagccac aatctactgc tccatacgga aacctcaggt ggtgccacca ccacaacaga 1620
atgatcttga gattcctgaa agtcctacct atgaaaattt cacctgaaag gaaaagcagc 1680
tgctgcctct ctcctgggac cgtggggttg gaaagtcagc tggacctcat ggggcctggg 1740
gctcgcagac agaagcacct cagaatttcc ttcagtgcct cagagatgcc tggatgtggc 1800
ccctcccccc ccttctcacc cttaaggact cccaaaccca ttaatagtt 1849

<210> 21
<211> 1427
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7682434CB1

<400> 21
cgccgcctct gccgcgatgc ccccccctgc gcccggggcc cggctccggc ttctcgccgc 60
cgccgccctg gccggcttgg ccgtcatcag ccgagggctg ctctcccaga gcctggagtt 120
caactctcct gccgacaact acacagtgtg tgaaggtgac aacgccaccc tcagctgctt 180
catcgacgag cacgtgaccc gcgtggcctg gctgaaccgc tccaacatcc tgtatgccgg 240
caatgaccgc tggaccagcg acccgcgggt gcggctgctc atcaacaccc ccgaggagtt 300
ctccatcctc atcaccgagg tggggctcgg cgacgagggc ctctacacct gctccttcca 360
gacccgccac cagccgtaca ccactcaggt ctacctcatt gtccacgtcc ctgcccgcat 420

tgtgaacatc tcgtcgcctg tgacggtgaa tgaggggggc aatgtgaacc tgctttgcct 480
ggccgtgggg cggccagagc ccacggtcac ctggagacag ctccgagacg gcttcacctc 540
ggagggagag atcctggaga tctctgacat ccagcggggc caggccgggg agtatgagtg 600
cgtgactcac aacggggtta actcggcgcc cgacagccgc cgcgtgctgg tcacagtcaa 660
ctatcctccg accatcacgg acgtgaccag cgcccgcacc gcgctgggcc gggccgccct 720
cctgcgctgc gaagccatgg cggttccccc cgcggatttc cagtggtaca aggatgacag 780
actgctgagc agcggcacgg ccgaaggcct gaaggtgcag acggagcgca cccgctcgat 840
gcttctcttt gccaacgtga gcgcccggca ttacggcaac tatacgtgtc gcgccgccaa 900
ccgactggga gcgtccagcg cctccatgcg gctcctgcgc ccaggatccc tggagaactc 960
agccccgagg cccccagggc tcctggccct cctctccgcc ctgggctggc tgtggtggag 1020
aatgtaggcg caacccagtg gagctcacct ccccctgcag ggggcctcag gccaagagtg 1080
agagaaacgg gggagcaaga gccgtgggtc tcgtgggggc agaagagctc tcggccacca 1140
aggaagaaga gagaggagaa gaggaggagg cagaggaaga aagatcttca gagaacccat 1200
cactgtgagg gataacgcaa aattatgcat ctttctacag ccattctcgc cacccgttca 1260
cgtttccgat tgtgacccac tcccgccacc ccataccccc ctctcttagc tcaggctgtc 1320
aactggcttg tgtgggtgtg ggtgtgtgag tgtgagcctg catgcatgtg taggtgtctg 1380
tgtctctgtt tgtgtgtgtg tgggggggtg ggctgggggа agggact 1427

<210> 22
<211> 1014
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2202389CB1

<400> 22
cacagatgcc actggggtag gtaaactgac ccaactctgc agcactcaga agacgaagca 60
aagccttcta cttgagcagt ttttccatca ctgatatgtg caggaaatga agacattgcc 120
tgccatgctt ggaactggga aattattttg ggtcttcttc ttaatcccat atctggacat 180
ctggaacatc catgggaaag aatcatgtga tgtacagctt tatataaaga gacaatctga 240
acactccatc ttagcaggag atccctttga actagaatgc cctgtgaaat actgtgctaa 300
caggcctcat gtgacttggt gcaagctcaa tggaacaaca tgtgtaaaac ttgaagatag 360
acaaacaagt tggaaggaag agaagaacat ttcatttttc attctacatt ttgaaccagt 420
gcttcctaat gacaatgggt cataccgctg ttctgcaaat tttcagtcta atctcattga 480
aagccactca acaactcttt atgtgacagg aaagcaaaat gaactctctg acacagcagg 540
aagggaaatt aacctggttg atgctcacct taagagtgag caaacagaag caagcaccag 600
gcaaaattcc caagtactgc tatcagaaac tggaatttat gataatgacc ctgacctttg 660
tttcaggatg caggaagggt ctgaagttta ttctaatcca tgcctggaag aaaacaaacc 720
aggcattgtt tatgcttccc tgaaccattc tgtcattgga ctgaactcaa gactggcaag 780
aaatgtaaaa gaagcaccaa cagaatatgc atccatatgt gtgaggagtt aagtctgttt 840
ctgactccaa cagggaccac tgaatgatca gcatgttgac atcattgtct gggctcaaca 900
ggatgtcaaa taatatttct caatttgaga atttttactt tagaaatgtt catgttagtg 960
cttgggtctt aagggtccat aggataaatg attaaaattt ctctcagaaa ctta 1014

<210> 23
<211> 3695
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7503597CB1

<400> 23

cccgcctgag gaagccgtgt gcctgggatg ccaagagcca gagaatggat cttctccgag 60
tggggacatt gctgacaatc ccggcttccc gaggcggcta agaacaggca gtttgtgtcg 120
gctggctgca gatacccaga ggcacaaaga gaccgaagcc acccggaggg acccacggac 180
ggacagatgg taggcgcgaa cccgagagga ccggcggagg ctgagcaccg agagccgcca 240
aggaagagaa actaaccaca gccaagttac cccgccggct ttccttcgct gcactaagga 300
atgaaaccct tccagctcga tctgctcttc gtctgcttct tcctcttcag tcaagagctg 360
ggcctccaga agagaggatg ctgtctggtg ctgggctaca tggccaagga caagtttcgg 420
agaatgaatg aaggccaagt ctattccttc agccagcagc cccaggacca ggtggtggtg 480
tcgggacagc cagtgacgct actttgcgcc atccccgaat acgatggctt cgttctgtgg 540
atcaaggacg gcttggctct gggtgtgggc agggacctct caagttaccc acagtacctg 600
gtggtaggga accacctgtc aggggagcac cacctgaaga tcctgagggc agagctgcaa 660
gacgatgcgg tgtacgagtg ccaggccatc caggccgcca tccgctcccg ccccgcacgc 720
ctcacagtcc tggtgccgcc tgatgacccc gtcatcctgg ggggcctgt gatcagcctg 780
cgtgcggggg accctctcaa cctcacctgc cacgcagaca atgccaagcc tgcagcctcc 840
atcatctggt tgcgaaaggg agaggtcatc aatggggcca cctactccaa gaccctgctt 900
cgggacggca agcgggagag catcgtcagc accctcttca tctcccctgg tgacgtggag 960
aatggccaga gcatcgtgtg tcgtgccacc aacaaagcca tccccggagg aaaggagacg 1020
tcggtcacca ttgacatcca gcaccctcca ctggtcaacc tctcggtgga gccacagcca 1080
gtgctggagg acaacgtcgt cactttccac tgctctgcaa aggccaaccc agctgtcacc 1140
cagtacaggt gggccaagcg gggccagatc atcaaggagg catctggaga ggtgtacagg 1200
accacagtgg actacacgta cttctcagag cccgtctcct gtgaggtgac caacgccctg 1260
ggcagcacca acctcagccg cacggttgac gtctactttg ggccccggat gaccacagaa 1320
ccccaatcct tgctcgtgga tctgggctct gatgccatct tcagctgcgc ctggaccggc 1380
aacccatccc tgaccatcgt ctggatgaag cggggctccg gagtggtcct gagcaatgag 1440
aagaccctga ccctcaaatc cgtgcgccag gaggacgcgg gcaagtacgt gtgccgggct 1500
gtggtgcccc gtgtgggagc cggggagaga gaggtgaccc tgaccgtcaa tggaccccc 1560
atcatctcca gcacccagac ccagcacgcc ctccacggcg agaagggcca gatcaagtgc 1620
ttcatccgga gcacgccgcc gccggaccgc atcgcctggt cctggaagga gaacgttctg 1680
gagtcgggca catcggggcg ctatacggtg gagaccatca gcaccgagga gggcgtcatc 1740
tccaccctga ccatcagcaa catcgtgcgg gccgacttcc agaccatcta caactgcacg 1800
gcctggaaca gcttcggctc cgacactgag atcatccggc tcaaggagca agagtctgtg 1860
ccgatggccg tcatcattgg ggtggccgta ggagctggtg tggccttcct cgtccttatg 1920
gcaaccatcg tggcgttctg ctgtgcccgt tcccagagaa atctcaaagg tgttgtgtca 1980
gccaaaaatg atatccgagt ggaaattgtc cacaaggaac cagcctctgg tcgggagggt 2040
gaggagcact ccaccatcaa gcagctgatg atggaccggg gtgaattcca gcaagactca 2100
gtcctgaaac agctggaggt cctcaaagaa gaggagaaag agtttcagaa cctgaaggac 2160
cccaccaatg gctactacag cgtcaacacc ttcaaagagc accactcaac cccgaccatc 2220
tccctctcca gctgccagcc cgacctgcgt cctgcgggca agcagcgtgt gcccacaggc 2280
atgtccttca ccaacatcta cagcaccctg agcggccagg gccgcctcta cgactacggg 2340
cagcggtttg tgctgggcat gggcagctcg tccatcgagc tttgtgagcg ggagttccag 2400
agaggctccc tcagcgacag cagctccttc ctggacacgc agtgtgacag cagcgtcagc 2460
agcagcggca agcaggatgg ctatgtgcag ttcgacaagg ccagcaaggc ttctgcttcc 2520
tcctcccacc actcccagtc ctcgtcccag aactctgacc ccagtcgacc cctgcagcgg 2580
cggatgcaga ctcacgtcta aggatcacac accgcgggtg gggacgggcc agggaagagg 2640
tcagggcacg ttctggttgt ccagggacga ggggtacttt gcagaggaca ccagaattgg 2700
ccacttccag gacagcctcc cagcgcctct gccactgcct tccttcgaag ctctgatcaa 2760
gcacaaatct gggtccccag gtgctgtgtg ccagaggtgg gcgggtgggg agacagacag 2820
aggctgcggc tgagtgcgct gtgcttagtg ctggacaccc gtgtccccgg ccctttcctg 2880
gaggcccctc taccacctgc tctgcccaca ggcacaagtg gcagctataa ctctgctttc 2940
atgaaactgc ggtccactct ctggtctctc tgtgggctct acccctcgct gaccagaagc 3000
tctacctacc cctgtgcctg tgctcccata cagccctggg gagaagggga tgacgtcttc 3060
ccagcactga gctgccccag aaaccccggc tccccactgc tgctcatagc ccatacccctg 3120
gaggctgaca agccagaaat ggccttggct aaaggagcct ctctctcacc aggctggccg 3180
ggagcccacc cccaatttgt ttggtgtttt gtgtccatac tcttgcagtt ctgtccttgg 3240
acttgatgcc gctgaactct gcggtgggac cggtccggtc agagcctggt gtactggggg 3300
gagggaggga ggagggagcc tgtgctgacg gagcacctcg ccgggtgtgc ccctcctggg 3360

ctgtgtgacc ccagcctccc cacccacctc ctgctttgtg tactcctccc ctccccctca 3420
gcacaatcgg agttcatata agaagtgcgg gagcttctct ggtcagggtt ctctgaacac 3480
ttatggagag agtgcttcct gggaagtgtg gcgtttgaag gggctggagg gcaggtcttt 3540
aagatggcga gactgccctt ctcagctgat aaacacaaga acggcgatcc tgtcttcagt 3600
aaggctccac gagaagagag gaagtatatc tacacctcaa ccctcctagt caccacctga 3660
aataaatgtt agggacacta ctccaaaaaa aaaaa 3695

<210> 24
<211> 2403
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7503603CB1

<400> 24
caggaataag tgacagtaag aatgacaagg gattaggact ggcttcctct tataaataat 60
aaaatccaaa gagaagtgac ttgagtctcc aggtttaaag gagagcaact agaagtcgtc 120
caaacacctg catctcataa ggagaagaaa agtccacctg gatcttgttt ctggactgag 180
atggatggag aggccacagt gaagcctgga gaacaaaagg aagtggtgag gagaggaaga 240
gaagtggact actccaggct cattgctggc actttaccac aatctcacgt tcttctttcg 300
cccttccaca aaaaagaccc catccgagat ggctgtggga gggctctctc ccctccagga 360
cccatctccg ggccatggga acactcaggc cttcctcgcc cctctgctgg cgggaggaga 420
gctcctttgc agctccaaat tcattgaagg gctcaaggct ggtgtcaggg gagcctggag 480
gagctgtcac catccagtgc cattatgccc cctcatctgt caacaggcac cagaggaagt 540
actggtgccg tctggggccc ccaagatgga tctgccagac cattgtgtcc accaaccagt 600
atactcacca tcgctatcgt gaccgtgtgg ccctcacaga ctttccacag agaggcttgt 660
ttgtggtgag gctgtcccaa ctgtccccgg atgacatcgg atgctacctc tgcggcattg 720
gaagtgaaaa caacatgctg ttcttaagca tgaatctgac catctctgca ggtcccgcca 780
gcaccctccc cacagccact ccagctgctg gggagctcac catgagatcc tatggaacag 840
cgtctccagt ggccaacaga tggaccccag gaaccaccca gaccttagga caggggacag 900
catgggacac agttgcttcc actccaggaa ccagcaagac tacagcttca gctgagggaa 960
gacgaacccc aggagcaacc aggccagcag ctccagggac aggcagctgg gcagagggtt 1020
ctgtcaaagc acctgctccg attccagaga gtccaccttc aaagagcaga agcatgtcca 1080
atacaacaga aggtgtttgg gagggcacca gaagctcggt gacaaacagg gctagagcca 1140
gcaaggacag gagggagatg acaactacca aggctgatag gccaagggag gacatagagg 1200
gggtcaggat agctcttgat gcagccaaaa aggtcctagg aaccattggg ccaccagctc 1260
tggtctcaga aactttggcc tgggaaatcc tcccacaagc aacgccagtt tctaagcaac 1320
aatctcaggg ttccattgga gaaacaactc cagctgcagg catgtggacc ttgggaactc 1380
cagctgcaga tgtgtggatc ttgggaactc cagctgcaga tgtgtggacc agcatggagg 1440
cagcatctgg ggaaggaagc gctgcagggg acctagatgc tgccactgga gacagaggtc 1500
cccaagcaac actgagccag accccggcag taggaccctg gggacccact ggcaaggagt 1560
cctccgtgaa gcgtactttt ccagaagatg aaagcagctc tcggaccctg gctcctgtct 1620
ctaccatgct ggccctgttt atgcttatgg ctctggttct attgcaaagg aagctctgga 1680
gaaggaggac ctctcaggag gcagaaaggg tcaccttaat tcagatgaca cattttctgg 1740
aagtgaaccc ccaagcagac cagctgcccc atgtggaaag aaagatgctc caggatgact 1800
ctcttcctgc tggggccagc ctgactgccc cagagagaga aatccaggac cctgagggac 1860
agagagatga actgctcagt taccatggga gaaggaccaa gatcaaaggc cttcaggacc 1920
ccagcctctt tccatcatcc ttcctccacc tgtgggaaga gaagctgatg cagccggtgc 1980
tccacccatg gaagaaaggc tggctgtcct tgggcccaag aaagtcaagc attatccacg 2040
tccaaaggtg acaagatgac tcaaaggaga cttcaagaac agtgtatgaa acactggaag 2100
aggtcaccta ggaaaagcat gaaatttcca ttcctgaatg tttgaaaata gaagaggctt 2160
ccaatcagtg tggaaagtga caaatcccct atcaacactc ccagcccttg ctgggggctc 2220
cttttctgac tactgttagc actcagcctc ccattcacat gtattatatt taagtgtacc 2280
agccttgcct tctcaagtag attctaagct cctttaaggc agtaattgca ttttatctgt 2340

ctcatgatgc ccccagagaa cttccaactc agtaggaacc catttaatac ctgtgtctga 2400
ttg 2403